Research Article
doi: 10.1590/01047760202330013317
vol(30), 2024

CERNE

A generalized mixed-effects height-diameter model for intensively managed *Pinus taeda* stands in Southern Brazil

Ximena Mendes de Oliveira[1*], Henrique Ferraço Scolforo[2], John Paul McTague[3], Mário Dobner Junior[4], José Roberto Soares Scolforo[5]

[1]Federal Rural University of Amazonia, Campus Parauapebas, Parauapebas, Pará, Brazil
[2]North Carolina State University, Raleigh, North Carolina, United States of America
[3]University of Georgia, Warnell School of Forestry and Natural Resources, Athens, Georgia, United States of America
[4]Florestal Gateados, Campo Belo do Sul, Santa Catarina, Brazil
[5]Federal University of Lavras, Campus Lavras, Department of Forest Science, Minas Gerais, Brazil

FOREST MANAGEMENT

ABSTRACT

Background: The insertion of stand variables allows for generalized mixed-effects height-diameter model, and the inclusion of random effects aids in model flexibility and application to new databases through calibration. This study aimed to develop a generalized mixed-effects model capable of accurately predicting the total height of *Pinus taeda* trees, subjected to different management regimes and at different ages.

Results: Experimental data between 4 and 27 years of age collected from stands of *Pinus taeda* located in the State of Santa Catarina were used. The experiment covers a combination of different planting densities and thinning intensities. The Lasso method was used to select the model's independent variables, and mixed modeling was implemented, with calibration using BLUP (best linear unbiased predictor). Dominant height and quadratic mean diameter were selected in different formulations by the Lasso method to compose the predictor variables of the model, along with diameter at 1.3 m (dbh). Random effects were entered into the intercept and inverse of dbh terms of the model.

Conclusion: The results show that the generalized model developed presents flexibility and can be applied to *Pinus taeda* stands in southern Brazil using the calibration of random effects, when necessary, by collecting a small sample of plot data.

Keywords: dominant height; quadratic mean diameter; thinning.

HIGHLIGHTS

The model is flexible to different management regimes, ages, and sociological positions.
The random effect was inserted in the intercept of the model.
The calibration can be applied considering the collection of a small sample of plot data.
The methodology can be applied to develop models for other species.

OLIVEIRA, X. M.; SCOLFORO, H. F.; TAGUE, J. P. MC.; DOBNER JÚNIOR, M.; SCOLFORO, J. R. S. A generalized mixed-effects height-diameter model for intensively managed *Pinus taeda* stands in southern Brazil. CERNE, v.30, 2024, e-103317, doi: 10.1590/01047760202430013317.

✤Corresponding author: ximena@ufra.edu.br
Received: August 1/2023 Accepted: August 19/2024

CERNE
ISSN 0104-7760
UNIVERSIDADE FEDERAL DE LAVRAS

INTRODUCTION

Height-diameter models are based on the relationship between the trees' total height and the diameter at 1.3 m (dbh). The degrade or lack-of-fit of this relationship negatively affects the models' development and can be caused by several factors, such as: poorly managed stands, where there are different heights for the same dbh (Batista et al., 2014); thinned stands, because they allow diameter growth to accelerate at the expense of height growth, and this trend becomes more obvious as thinning intensity increases (Deng et al., 2019); stands at advanced ages, because with increasing age, height growth attenuates while the dbh continues to increase (Marziliano et al., 2019).

The degrade of the height-diameter relationship can be countered by the use of plot-specific or local height-diameter models (Ribeiro et al., 2010). Generalized models are a viable alternative for solving this issue and avoiding the necessity for a large number of models developed for specific stands or groups of stands (Adame et al., 2008; Bronisz and Mehtätalo, 2020). A generalized height-diameter model is also a must for a simulated yield.

Using stand variables is the path to generalization. Variable selection methods can identify which variables should be included in the model. The Lasso method (Least Absolute Shrinkage and Selection Operator), proposed by Tibshirani (1996), consists of a semi-parametric technique that equally penalizes all independent variables of linear models through Machine Learning, aiming to minimize the sum of squares of the errors, provided that the absolute sum of the coefficient values is less than a penalty parameter, represented by "λ" (lambda) (Altoé, 2017). Fiorentin et al. (2019) and Farjat et al. (2015) used the Lasso method as one of the variable selection strategies to develop generalized linear models.

However, the mere inclusion of variables does not guarantee that a generalized model will maintain accuracy when applied to a situation completely different from that considered in the fit. One solution to this problem is mixed-effects models, which receive this name for having fixed and random effects in their composition. The fixed effects are associated with the population, while the random effects are associated with specific groups, such as sample units or treatments. By associating random effects to observations of the same rank level, mixed-effects models contain a covariance structure induced by the data clustering, allowing the correction of any erroneous trend of the generalized model (Pinheiro and Bates, 2013; Gómez-García et al., 2015).

The development of generalized models for estimating total height in intensively managed pine stands in Brazil is of great relevance. Pine has high economic value and is the second most planted genus in the country (IBÁ, 2023). Stands are managed with a wide variety of initial densities and thinning regimes (David et al., 2018), aiming to generate products that serve different sectors and provide greater profitability (Dobner Jr. and Quadros, 2019). Therefore, simulating the destination of the wood for the different markets requires accurate model performance for a variety of management regimes.

The introductory paragraphs contained here are in support of several hypotheses regarding the height-diameter relationship: I) Generalized models that incorporate stand-level attributes of size, site quality, and competition, or environmental effects improve regional performance; II) Multicollinearity among the stand-level and environmental variables can be severe, and LASSO represents an efficient methodology for variable selection; III) When height-diameter models are constructed with random effects, they can be efficiently calibrated and localized with a small sample size of a new population.

This study sought to perform an approach combining a selection of relevant variables for generalization using the Lasso method with mixed modeling for calibration of stands. The overall aim was to develop a generalized mixed-effects height-diameter model used in *Pinus taeda* stands conducted under different management options in southern Brazil.

MATERIAL AND METHODS

Study area

The study area is located in the municipality of Campo Belo do Sul, State of Santa Catarina, southern Brazil (lat. 27°59'33"S, long. 50°54'16"W). According to Alvares et al. (2013), the Köppen climate classification is a humid temperate climate with temperate summer (Cfb), characterized by the absence of a dry season, a maximum average temperature of 22°C in the warmest month, and an average temperature below 10°C in at least four months throughout the year. Frosts are frequent, ranging from two to 29 events per year (Braga et al., 2021). The soils in the region are classified as Litholic Neossol, Brune Latossol, and Haplic Nitossol (Oliveira et al., 2014).

Characterization of the experiment

The data come from *Pinus taeda* stands, conducted in a 3x3 factorial experiment with two repetitions (plots), implemented in 1984. The plots were installed with a size of approximately 2000 m^2 with a functional measurement area of nearly 1000 m^2. The two factors of the experiment are: planting density (2500, 1250, and 625 $trees.ha^{-1}$) and thinning (without thinning (WT), representing the control treatment; moderate thinning (MT); and heavy thinning (HT)), totaling eighteen plots and nine treatments (Table 1).

The treatments submitted to thinning (letters d - i) were conducted using the methodology known as "crown thinning", described in detail by Dobner Jr. (2015) and Oliveira et al. (2024). Pruning was performed only in the thinning treatments, with two lifts, as follows: pruning all trees up to 2.5 m at five years of age and pruning only the selected potential trees up to 6 m at seven years.

Table 1: Description of the factorial experiment with nine treatments (letters a - i) and percentage of tree removal in each thinning performed.

Treatments			
Factor: thinning	Factor: initial density		
	2500 trees.ha^{-1}	1250 trees.ha^{-1}	625 trees.ha^{-1}
Without thinning (WT)	**a**) WT_2500	**b**) WT_1250	**c**) WT_625
Moderate thinning (MT)	**d**) MT_2500	**e**) MT_1250	**f**) MT_625
Heavy thinning (HT)	**g**) HT_2500	**h**) HT_1250	**i**) HT_625

Removal percentage of the number of trees					
Treatment	Plot	Age (years)			
		5	8	12	27
d) MT_2500	1	18 %	16 %	11 %	6 %
	2	16 %	14 %	13 %	5 %
e) MT_1250	1	-	24 %	22 %	11 %
	2	-	30 %	14 %	8 %
f) MT_625	1	-	26 %	19 %	11 %
	2	-	31 %	25 %	8 %
g) HT_2500	1	30 %	20 %	17 %	8 %
	2	35 %	15 %	9 %	12 %
h) HT_1250	1	-	58 %	11 %	10 %
	2	-	54 %	10 %	11 %
i) HT_625	1	-	66 %	-	9 %
	2	-	67 %	-	7 %

DATABASE

The diameter data in cm at 1.3 m (dbh) and total height in m (h) were measured when the trees were 4, 10, 12, and 27 years old. In all, 1,711 height-dbh pair observations were collected. The dominant height followed the concept of Assmann (1970), considering the 100 trees with the highest dbh in the hectare. The variability of this information at the tree level and the other stand variables (Table 2) ensured a representative database for the different management options applied.

The variable uniformity index (ui) was obtained according to Equation (1), aiming to represent how much the plot variability alters the diameter distribution over time so that the greater the diameter range, the lower the uniformity value obtained in the index (Oliveira et al., 2022).

$$ui_{jk} = \frac{1}{(dbh\_p63_{jk} - dbh\_p10_{jk})} \quad (1)$$

j: plots ranging from 1 to 18; k: age (years) ranging from 4 to 27; ui_{jk}: uniformity index (cm^{-1}) of the j-th plot at age k; dap_p63_{jk}: dbh (cm) located at the 63rd percentile of the j-th plot at age k; dbh_p10_{jk}: dbh (cm) located at the 10th percentile of the j-th plot at age k.

Additionally, the data collected allowed calculating the average thinning index per treatment over the years (Equation 2).

$$ti_{jk} = \frac{(in_j - rn_{jk})}{in_j} \quad (2)$$

ti_{jk}: thinning index of the j-th plot at age k; in_j: initial number of trees in the j-th plot, rn_{jk}: removed number of trees from the j-th plot at age k.

Proposed approach to height-diameter modeling

In generalized height-diameter models, the variable h is modeled based on dbh and other stand variables (Scolforo et al., 2019; Bronisz and Mehtätalo, 2020). The full base linear model (Equation 3) was considered in this study. From this model, modeling was performed following two steps, the first to select the variables and the second to add a random effect in the model.

$$\begin{aligned} \ln(h) = \beta_0 + \beta_1\left(\frac{1}{dbh}\right) + \beta_2 \ln(H) + \beta_3 \ln\left(\frac{dq}{dbh}\right) + \\ +\beta_4\left(\frac{1}{age.dbh}\right) + \beta_5\left(\frac{1}{age}\right) + \beta_6 \ln(B) + \beta_7 \ln(N) + \\ +\beta_8 \ln(dq) + \beta_9 (ui) + \beta_{10} \ln(ui) + \beta_{11}(ti) + \beta_{12} \ln(ti) + \varepsilon \end{aligned} \quad (3)$$

$\beta_{i's}$: fixed coefficients of the model. Other variables have been described previously.

Table 2: Descriptive statistics of dendrometric variables in *Pinus taeda* stands.

Statistics	Variables								
	h (m)	dbh (cm)	dq (cm)	H (m)	B (m^2. ha^{-1})	N (árv.ha^{-1})	Age (years)	ti	ui (cm^{-1})
Min.	2.70	4.2	8.12	5.03	1.32	179.93	4.00	0.31	0.07
Max.	36.10	67.4	53.89	34.49	93.94	2481.19	27.00	1.00	0.55
Med.	13.74	21.85	21.73	14.35	30.65	1153.13	10.86	0.82	0.21
SD	9.06	12.79	12.06	8.96	24.50	695.62	7.89	0.25	0.12

dq (cm): quadratic diameter of the plot at each age; H (m): dominant height, obtained by the Assmann method (1970) of the plot at each age, that considers the average height of the 100 largest diameter stems/ha; B (m^2.ha^{-1}): basal area of the plot at each age; N (trees.ha-1): number of trees.ha^{-1} of the plot at each age; ti: thinning index of the plot at each age; ui: uniformity index of the plot at each age. Other variables have been described previously.

Step 1: Variable Selection

Step 1 is essential for selecting independent variables for the model (Equation 3). The Lasso method was employed for variable selection. Lasso is a semi-parametric method that penalizes the coefficients and selects the explanatory variables in the model. The penalty strength is given by lambda (λ) so that the higher the λ, more coefficients reduce to zero, and fewer predictors are selected in the model (Tibshirani, 1996). In this study was considered the lambda at one deviation.

Step 2: Adding the random effect in the generalized model

Following the variable selection, the inclusion of the random effect for plots was introduced for the intercept and other predictors, aiming to allow more flexibility to the model with the mixed-effects approach. The linear mixed-effects model can be represented in a matrix form in general (Pinheiro and Bates, 2013):

$$y = X\beta + Zb + \varepsilon, b \sim N(0,\hat{D}) \text{ and } \varepsilon \sim N(0,\hat{R})$$

y: vector with dependent variables observed in the data; X: fixed-effects design matrix; $\hat{\beta}$: vector with fixed-effect parameters; Z: random-effects incidence matrix; b: vector with random-effect parameters; ε: vector with random errors; N: normal distribution; $\hat{D}$: matrix of variances and covariances of the random parameters, $\hat{R}$: matrix of variances and covariances of the random errors.

Analysis of the proposed approach

The model performance was assessed with root mean square error (RMSE, Equation 4), mean absolute error (MAE, Equation 5), and mean error (T, Equation 6) of the untransformed h variable. In addition, graphs were generated considering the mean residuals in ten standardized dbh classes, the 95% confidence interval of the individual observations (mean ± 1.96 standardized diameter), and the confidence interval for the class mean (mean ± 1.96 standardized error). The standardized dbh was calculated as the difference between the tree's dbh and the plot mean dbh, divided by the standard deviation of the plot's dbh.

$$RMSE(m) = \sqrt{\frac{\sum_{i=1}^{n}(O-P)^2}{n}} \quad (4)$$

$$MAE(\%) = \frac{\sum_{i=1}^{n}\left(\frac{|O-P|}{O}\right)}{n}100 \quad (5)$$

$$T(\%) = \frac{\sum_{i=1}^{n}\left(\frac{O-P}{O}\right)}{n}100 \quad (6)$$

n: number of trees in the sample ranging from 1 to n; O: observed values of h (m), P: predicted values of h (m) without the logarithmic correction factor.

In addition to the general evaluation of the developed model, residual plots and model performance statistics were obtained by thinning treatment, ages, sites, and sociological positions to evaluate the model's flexibility in different situations. The behavior of total height over time for the different site conditions, sociological position, and treatment (initial density vs. thinning) was analyzed graphically using the adjusted model.

Processing was performed in R software (R Core Team, 2023). Graphs were generated using the ggplot2 (Wickham, 2022) and lmfor (Mehtätalo, 2022) packages. The glmnet package (Friedman et al., 2023) was used for variable selection with the Lasso method. The nlme package (Pinheiro et al., 2023) was used to adjust the linear mixed models.

Example of model calibration

The main goal of developing a generalized mixed-effects model is to enable its application or extrapolation to new data, outside of the fitting database. The model calibration is needed in this situation and consists in obtaining estimated values for the random parameters. For this purpose, a small data sample is sufficient and different methodologies can be applied to select trees to compose this reduced sample (Bronisz and Mehtätalo, 2020).

This study used BLUP (best linear unbiased predictor) (7) to provide an example on how to estimate the random parameters. The predictor variance is represented in (8) (Lappi, 1991).

$$\hat{b} = (Z'\hat{R}^{-1}Z + \hat{D}^{-1})^{-1}Z'\hat{R}^{-1}(y - X\beta) \quad (7)$$

$$var(\hat{b} - b) = (Z'\hat{R}^{-1}Z + \hat{D}^{-1})^{-1} \quad (8)$$

The calibration was shown on plot 1 of treatment (e) MT_1250. To do so, the data from this plot was removed from the database, and the model was fit to the reduced dataset. The subset of this plot used in the model calibration was composed of 32 trees (8 per age) by the extreme tree method, that is, selecting the four thinnest and four thickest trees at each age (Bronisz and Mehtätalo, 2020).

RESULTS

Selection of model predictor variables

Three variables and the intercept of Equation (3) were selected by the Lasso method, considering λ at one deviation: 1/dbh, ln(dq/dbh), and ln(H). The lambda at one deviation was selected as it was less restrictive in the selection of variables and generated consistent results. Although all tested variables correlate with h (all correlations were significant, considering α=0.05), the selected predictor variables are highly correlated with the other independent variables.

The λ penalty considered a λ range of 0.0032 to 0.7078, and the λ at one deviation corresponded to λ=0.0108. When λ = 0.0108, nine coefficients from Equation (3) had an estimated value of zero and were not selected to compose the model. The Lasso method was used to identify variables that could remain in the model.

Generalized mixed-effects model

Upon declaring the intercept and the inverse dbh term to be random variables, Model (Equation 9) resulted from the mixed modeling combined with the variables selected by Lasso:

$$\ln(h) = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})\left(\frac{1}{dbh}\right) + \beta_2 \ln(H) + \beta_3 \ln\left(\frac{dq}{dbh}\right) + \varepsilon \quad (9)$$

$b_{f's}$: Random effects coefficients for plots. Other variables have been described previously.

Figure 1a shows the 1:1 plot with the adjustment performance statistics of the generalized mixed-effects model (9) and Figure 1b shows a homogeneous distribution of Pearson's residuals along the standardized dbh values.

Estimates, standard errors, and p-value of the fixed effects are presented in Table 3. Random and residual parameter variances and covariances and the correlation between random parameters are presented in Table 4.

Table 3: Estimated parameters, standard errors, and p-value of the generalized mixed-effects model (9).

Fixed parameters	Estimated Value	Standardized Error	p-value
β_0	0.2584	0.0271	~0.0000
β_1	-2.5271	0.2179	~0.0000
β_2	0.9340	0.0072	~0.0000
β_3	-0.1540	0.0122	~0.0000

Table 4: Random and residual parameter variances and covariances and correlation among the random parameters of the generalized mixed-effects model (9).

	Variances and covariances			Correlation		
	b_{0j}	b_{1j}	ε_{ijk}	b_{0j}	b_{1j}	ε_{ijk}
b_{0j}	0.0005			1		
b_{1j}	-0.0126	0.3757		-0.92	1	
ε_{ijk}	0	0	0.0040	0	0	1

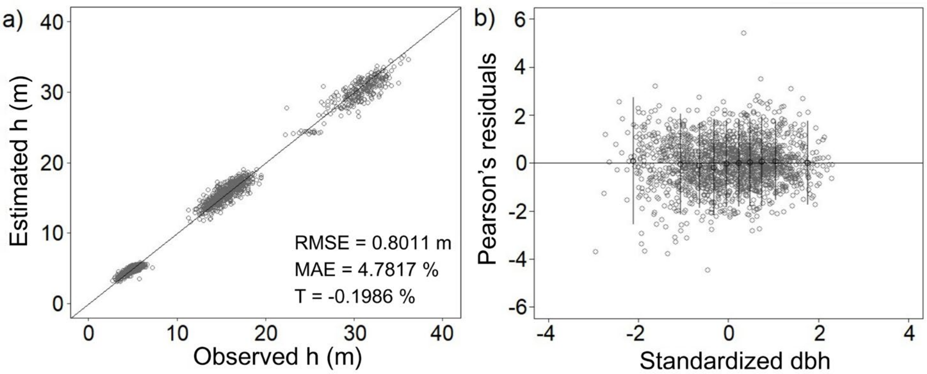

Figure 1: a - 1:1 ratio (observed h: estimated h) and performance statistics of the generalized mixed-effects model (9); b - Residuals plot of the generalized mixed-effects model (9).

Prediction of the generalized mixed-effects model in different situations

After the general analysis, model (9) was evaluated in different situations. Pearson's residuals remained unbiased along the standardized dbh variable concerning each density x thinning treatment, and the maximum values of RMSE, MAE, and T were 1.0096 m, 5.5177 %, and -0.7051%, respectively (Figure 2).

Pearson's residuals also remained unbiased along standardized dq/dbh variable regarding different ages, sites, and sociological positions (Figure 3). The stratification of sociological positions was performed, according to dq/dbh, into three classes: suppressed (less than 1 minus one standard deviation of dq/dbh), dominant (greater than 1 plus one standard deviation of dq/dbh), and intermediate (range between suppressed and dominant classes).

Total height estimation for different site conditions, sociological position, and management practices

The behavior of estimated total height over time was analyzed for different site conditions, sociological position, and management options using the fitted model (Figure 4). The estimates for the two sites showed little difference. The estimates for the sociological positions showed differences in height among the different classes. At initial densities of 2500 and 625 trees ha^{-1}, the treatment without thinning showed a trend slightly different from the others; however, at an initial density of 1250 trees ha^{-1}, this distinction was not observed.

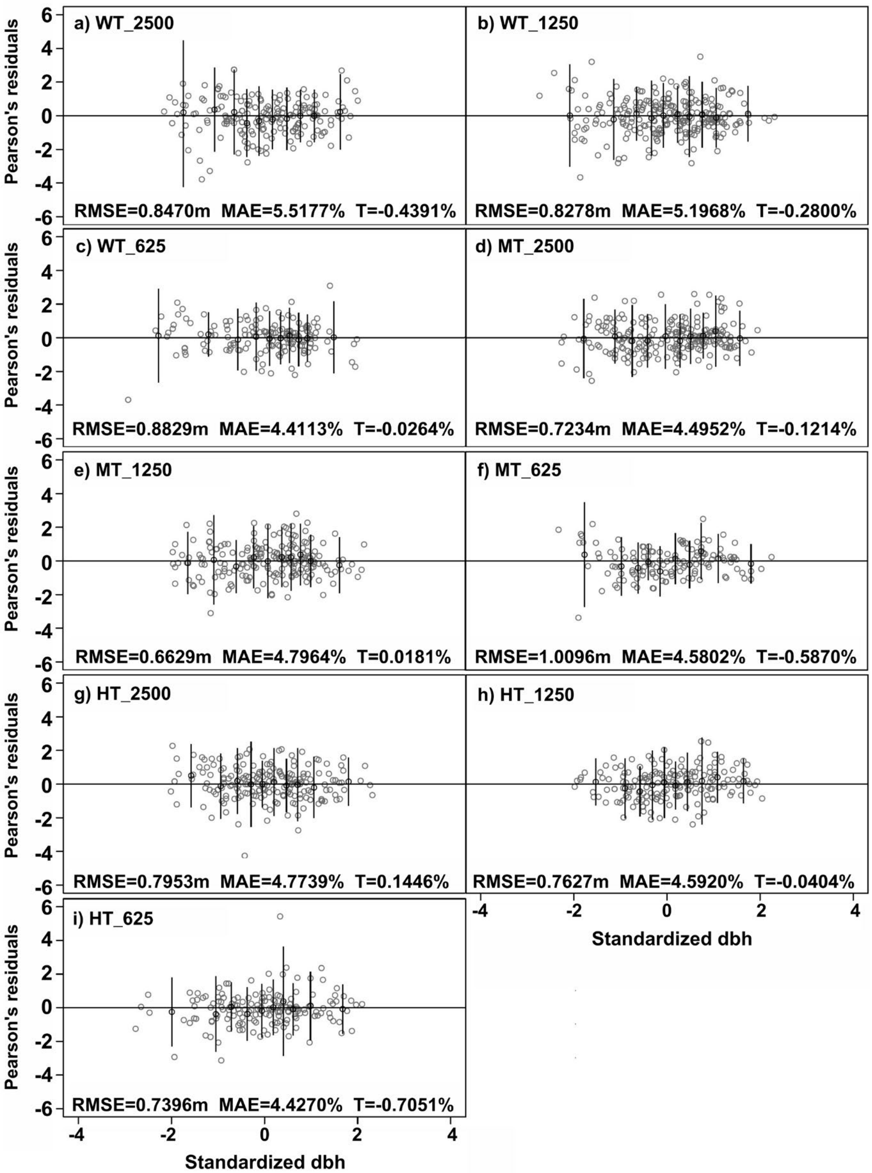

Figure 2: Residuals and performance statistics of the generalized mixed-effects model (9) in the different density x thinning treatments.

Example of model calibration

The fixed parameter estimates with the reduced dataset were β_0 = 0.2463; β_1 = -2.5008; β_2 = 0.9378, and β_3 = -0.1561. Next, calibration was performed using a subsample with 32 trees selected by the extreme tree methodology. To estimate the random plot parameters with BLUP (4), the matrices with random-effects incidence (Z), with variances and covariances of the random errors ($\hat{R}$), with variances and covariances of the random parameters ($\hat{D}$) and with fixed-effects design (X) and the vectors with dependent variables observed in the data (y) and with fixed-effect parameters (B) vectors were considered:

$$Z = \begin{bmatrix} 1 & 1/dbh_1 \\ \vdots & \vdots \\ 1 & 1/dbh_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0.2128 \\ \vdots & \vdots \\ 1 & 0.0192 \end{bmatrix}$$

$$\hat{R} = \sigma_\varepsilon^2 I = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & & \\ & \ddots & \\ & & \sigma_{\varepsilon_{32}}^2 \end{bmatrix} = \begin{bmatrix} 0.0041 & & \\ & \ddots & \\ & & 0.0041 \end{bmatrix}$$

$$\hat{D} = \begin{bmatrix} \sigma_{b_{0j}}^2 & \sigma_{b_{1j}}\sigma_{b_{0j}} \\ \sigma_{b_{0j}}\sigma_{b_{1j}} & \sigma_{b_{1j}}^2 \end{bmatrix} = \begin{bmatrix} 0.0005 & -0.0127 \\ -0.0127 & 0.3735 \end{bmatrix}$$

$$y = \begin{bmatrix} \ln(h_1) \\ \vdots \\ \ln(h_{32}) \end{bmatrix} = \begin{bmatrix} 1.1314 \\ \vdots \\ 3.4904 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1/dbh_1 & \ln(H_1) & \ln(dq_1/dbh_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1/dbh_{32} & \ln(H_4) & \ln(dq_4/dbh_{32}) \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 0.2128 & 1.6174 & 0.5962 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0.0192 & 3.4673 & -0.3059 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0.2463 \\ \vdots \\ -0.1561 \end{bmatrix}$$

After obtaining these data, the random effects were estimated, according to $\hat{b} = (Z'\hat{R}^{-1}Z + \hat{D}^{-1})^{-1} Z'\hat{R}^{-1}(y - X\beta)$ and the final calibrated equation was (10):

$$\ln(h) = (0.2463 - 0.0047) + (-2.5008 + 0.1488)\left(\frac{1}{dbh}\right) + 0.9378 * \ln(H) - 0.1561 * \ln\left(\frac{dq}{dbh}\right) \quad (10)$$

The estimated heights, considering only the fixed-effects parameters and considering the mixed effects of the calibrated equation (10), presented different results regarding the performance statistics. The equation with fixed effects presented RMSE = 0.7461 m, MAE = 5.2552 %, and T = 1.4889 % while the equation with mixed effects presented RMSE = 0.7392 m, MAE = 5.0799 %, and T = 1.2922 %.

DISCUSSION

Accurate estimates of total tree heights play an important role in forest stand management and monitoring (Batista et al., 2014). The data used in this study cover nine combinations between initial densities and thinning intensities at ages ranging from 4 - 27 years in *Pinus taeda* stands located in southern Brazil. The broad conditions compose a solid base to represent different management alternatives that can be used in plantations of this species (David et al., 2018; Huss and Dobner Jr., 2020), resulting in multiple products, such as pulp and paper industry, sawtimber, and panels (IBÁ, 2023). This study developed a generalized mixed-effects model for *Pinus taeda* trees that showed flexibility in different management regimes in southern Brazil. The model described the height-diameter relationship in stands with different thinning intensities; moreover, it provided accurate estimates at different sites, ages, and sociological positions.

Knowledge of the factors that affect the dependent variable has great relevance in gaining flexibility of the models, enabling generalization (Oliveira et al., 2024). Barros et al. (2002) developed a generalized linear fixed-effects height-diameter model, selecting stand variables using the Stepwise method. Chong and Jun (2005) claim that, although the Stepwise method is often used because of its simplicity, it often performs poorly when multicollinearity exists between variables. In that study, the authors found that Lasso showed, on average, lower RMSE compared to the Stepwise method. Ng et al. (2020) published a study defending the Lasso approach for reducing the number of covariates in a model. The authors employed the method to discriminate which variables should be in the model, enabling exploratory research. As in the present study, Ng et al. (2020) did not use the parameters estimated by the Lasso method as they considered the method only for variable selection.

Farjat et al. (2015) used the Lasso method to select climatic variables form modeling height growth *Pinus taeda* in southeastern United States, while Altoé (2017) employed Lasso to model the carbon stock of forest fragments in the State of Minas Gerais, Brazil. In addition to the stand variables commonly used in generalized models (Scolforo et al., 2019), variables representing thinning intensity and plot uniformity over time were tested. However, they were not selected by Lasso, indicating that the stand variables could represent these conditions.

Similar to this study, Vargas-Larreta et al. (2009) also considered quadratic mean diameter and dominant height to generalize a height-diameter model with pine data in Mexico. Saud et al. (2016) considered the inclusion of quadratic mean diameter and relative spacing index to enhance height–diameter models. Raptis et al. (2021) considered the inclusion of dominant height and dominant diameter. Bronisz and Mehtätalo (2020) did not consider the inclusion of dominant height but inspected quadratic mean diameter and basal area. Ferraz Filho et al. (2018) found that the dominant height is sensitive to the management regime adopted. Moreover, this variable can explain possible climatic variations between stands (Scolforo et al., 2013).

Although some studies adopt non-linear modeling to express the relationship between dbh and h (Adame et al., 2008), linear models have many advantages due to their simplicity and are widely used in simple and generalized forms (Silva et al., 2016; Fortin et al., 2019). Mehtätalo et al. (2015) note that a non-linear model with more than 2 parameters suffers from problems in model convergence and non-positive definite random-effect variance-covariance matrix. A linear model formulation of height greatly facilitates the procedure of calibrating localized models.

Including random effects of some variables in a model results in mixed-effects (fixed and random). While fixed effects are associated with the population, random effects are associated with specific groups, increasing the model's flexibility (Pinheiro and Bates, 2013). Ferraz Filho et al. (2018) developed two linear mixed models for eucalypt stands under different management conditions, adding the age and dominant height variables and random intercepts for plots and trees within plots. Lynch et al. (2005) did not include additional variables in the linear, log transformed h model as a function of dbh but assigned mixed effects to the model parameters and obtained a model with better accuracy and flexibility. Strimbu et al. (2018) discuss several correction factors for transformed dependent variable models, including the common factor for linear logarithmic models, $e^{RMSE^2/2}$. However, it should be clear that the correction factor is virtually immaterial, as the correction factor for RMSE = 0.1 is 1.00501.

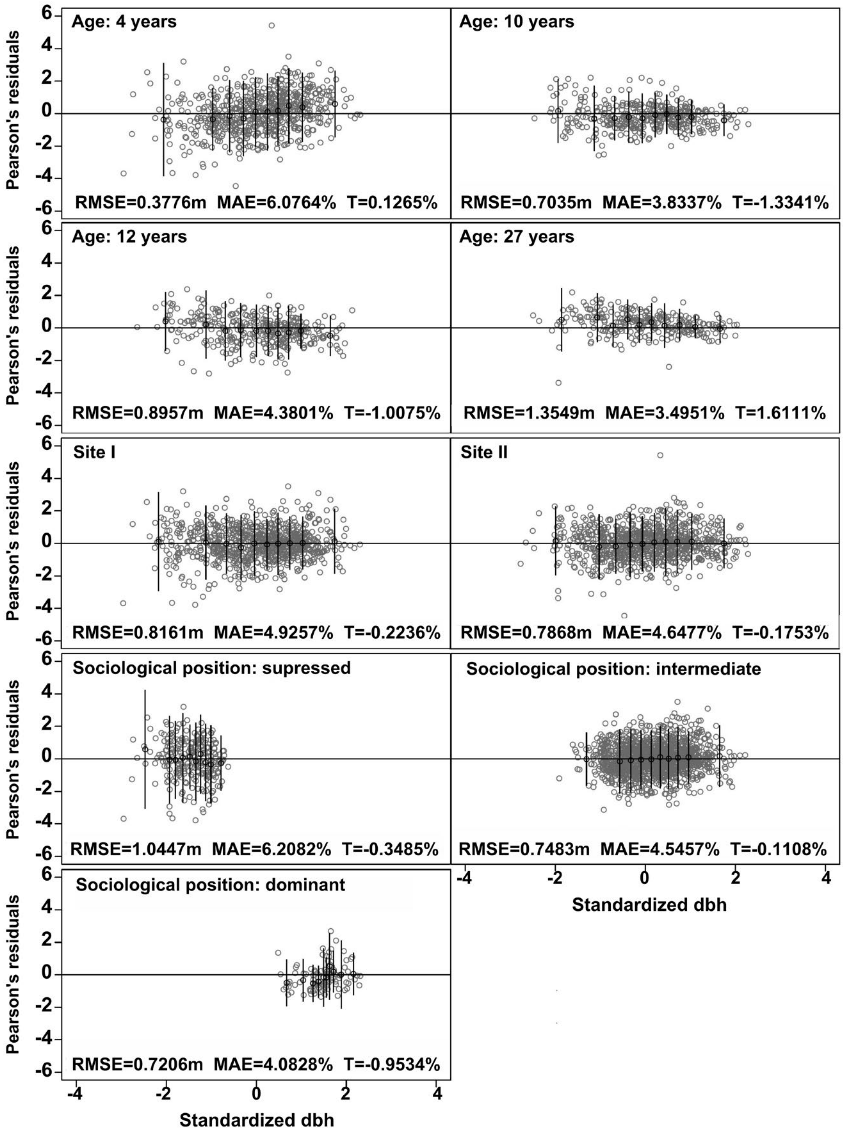

Figure 3: Residuals and performance statistics of the generalized mixed-effects model (9) in the different ages, sites, and sociological positions.

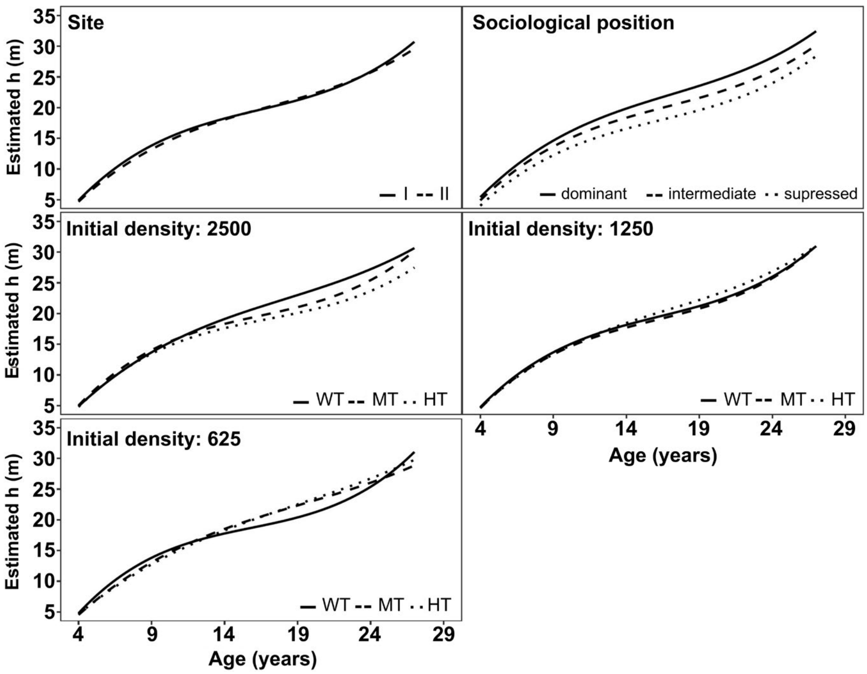

Figure 4: Total height estimates from the generalized mixed-effects model (9) at the different sites, sociological positions, and management treatments.

Any application should be analyzed in different situations to verify a model's flexibility. Ferraz Filho et al. (2018) verified the behavior of the deviation regarding different ages and dominant height. The model developed by these authors showed flexibility and could be used in different situations (different planting densities, thinning regimes, fertilization, and eucalyptus clones), avoiding a large number of models developed for specific situations, as in the study by Ribeiro et al. (2010) who divided the data into 14 treatments according to age and region.

Assigning mixed effects to a model allows for its use on new data with the re-estimation of the random parameters from an unbiased predictor through a process called calibration. A subsample of data is used for calibration. This process ensures that internal variations are considered correctly and well-grounded (Adame et al., 2008). Castedo Dorado et al. (2006) used dbh and h measurements of three randomly selected trees in the plot to re-estimate the random parameters. Raptis et al. (2021) proposed using measurements of two random trees. Bronisz and Mehtätalo (2020) tested different tree selection methods at different numbers and found that the tree selection method at extreme dbh values (smallest and largest) was the most effective; moreover, with an increasing number of trees, the RMSE decreases. In this study, the extreme tree method was considered, and through preliminary tests, it was verified that the model's performance statistics were more stable from eight trees per plot at each age.

Lynch et al. (2005) verified a significant improvement in total height estimates after model calibration. In this study, calibration also improved the accuracy of the estimates furnished in the example. Furthermore, it consists of a simple application method, as demonstrated in this study.

Future studies using the model developed in this study can be conducted in pine stands submitted to different forest management regimes in southern Brazil. For atypical situations, calibration can be implemented, improving the accuracy of estimates by collecting a small data sample.

CONCLUSION

A generalized linear mixed-effects height-diameter model was developed in this study. Three independent variables were selected by the Lasso method, considering the penalty of λ at one deviation. The mixed effect was considered in the intercept and inverse dbh variable of the model. The generalized linear mixed-effects model showed flexible behavior and can be applied to *Pinus taeda* stands in southern Brazil with different management regimes, ages, sites, and sociological positions. For situations where the plots are outside the data range used in this study, the calibration can be applied as shown, considering the collection of a small sample of plot data.

AUTHORSHIP CONTRIBUTION

Project Idea: XMO; JPM; JRSS

Funding: JRSS

Database: MDJ

Processing: XMO; JPM; JRSS

Analysis: XMO; JPM; MDJ; JRSS; HFS

Writing: XMO; JPM; JRSS; HFS

Review: XMO; JPM; MDJ; JRSS; HFS

ACKNOWLEDGMENTS

The authors are especially grateful to the private enterprise Florestal Gateados, Forest Management Laboratory (LEMAF) at the Federal University of Lavras (UFLA) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

ADAME, P.; DEL RÍO, M.; CAÑELLAS, I. A mixed nonlinear height–diameter model for Pyrenean oak (*Quercus pyrenaica* Willd). Forest Ecology and Management, v. 256, n.1-2, p. 88-98, 2008.

ALTOÉ, T. F. Modelagem e espacialização do estoque de carbono em fragmentos nativos de Minas Gerais em função de fatores edafoclimáticos. 2017. 130 p. PhD Thesis Universidade Federal de Lavras, Lavras.

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.

ASSMANN, E. The principles of forest yield study. Oxford: Pergamon Press, 1970. 506 p.

BARROS, D. A.; MACHADO, S. A.; ACERBI JÚNIOR, F. W.; SCOLFORO, J. R. S. Comportamento de modelos hipsométricos tradicionais e genéricos para plantações de *Pinus oocarpa* em diferentes tratamentos. Pesquisa Florestal Brasileira, v. 45, p. 3–28, 2002.

BATISTA, J. L. F.; COUTO, H. T. Z.; SILVA FILHO, D. F. Quantificação de recursos florestais: árvores, arvoredos e florestas. São Paulo: Oficina de Textos. 384 p. 2014.

BRAGA, G. B.; IMBUZEIRO, H. M. A.; PIRES, G. F. Frost Risk and Rural Insurance in Brazil. Revista Brasileira de Meteorologia, v. 36, n. 4, p. 703-711, 2021.

BRONISZ, K., MEHTÄTALO, L. Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. Forest Ecology and Management, v. 460, 117901, 2020.

CASTEDO DORADO, F.; DIÉGUEZ-ARANDA, U.; ANTA, M. B.; et al. A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. Forest Ecology and Management, v.229, n.1–3, p. 202–213, 2006.

CHONG, I. G.; JUN, C. H. Performance of some variable selection methods when multicollinearity is present. Chemometrics Intelligent Laboratory Systems, v.78, n.1-2, p. 103-112, 2005.

DAVID, H. C.; PÉLLICO NETTO, S.; ARCE, J. E.; et al. Intensive management for optimizing the production of high-value logs of pine forests in southern Brazil. Ciência Florestal, v.28, n.3, p.1303-1316, 2018.

DENG C.; ZHANG, S.; LU, Y.; et al. Thinning Effects on the Tree Height – Diameter Allometry of Masson Pine (*Pinus massoniana* Lamb). Forests, v.10, n.12, p.1-15, 2019.

DOBNER JR. M.; QUADROS, D. S. Economic performance of loblolly pine stands in southern Brazil as a result of different crown thinning intensities. Revista Árvore, v. 43, n. 2, p. 1-11, 2019.

DOBNER JR.; M. Desbaste pelo alto: uma alternativa rentável para povoamentos de *Pinus taeda* no Sul do Brasil. Santa Catarina State, Brazil. 28 p, 2015. Available at: https://engenhariaflorestal.paginas.ufsc.br/files/2015/09/Cartilha_desbaste_Dobner-e-Huss.pdf

FARJAT, A. E.; ISIK, F.; REICH, B. J.; et al. Modeling Climate Change Effects on the Height Growth of Loblolly Pine. Forest Science, v. 61, n. 4, p. 703–715, 2015.

FERRAZ FILHO, A. C.; MOLA-YUDEGO, B.; RIBEIRO, A.; et al. Height-diameter models for Eucalyptus sp. plantations in Brazil. Cerne, v. 24, n.1, p. 9–17, 2018.

FIORENTIN, L. D.; BONAT, W. H.; PELISSARI, A. L.; et al. Generalized linear models for tree survival in loblolly pine plantations. Cerne, v. 25, n. 4, p. 347–356, 2019.

FORTIN, M.; COUWENBERGHE, R. V.; PEREZ, V.; et al. Evidence of climate effects on the height-diameter relationships of tree species. Annals of Forest Science, v. 76, n. 1, 2019.

FRIEDMAN J, HASTIE T, TIBSHIRANI R. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 4.1-8. 2023. Available at: http://CRAN.R-project.org/package=glmnet

GÓMEZ-GARCÍA, E.; FONSECA, T. F.; CRECENTE-CAMPO, F.; et al. Height-diameter models for maritime pine in Portugal: a comparison of basic, generalized and mixed-effects models. iForest - Biogeosciences and Forestry, v. 9, n.1-2, p. 72-78, 2015.

HUSS, J.; DOBNER JR.; M. Einfluss von Ausgangspflanzendichte und Durchforstung auf das Wachstum von Weihrauchkiefern (*Pinus taeda*) in Südbrasilien. Allgemeine Forst-und Jagdzeitung, v.190, n.1-21, 2020.

IBÁ (Indústria Brasileira de Árvores), 2023. Annual report of IBÁ 2023, base year 2022. Available at: https://www.iba.org/datafiles/publicacoes/relatorios/relatorio-anual-iba2023-r.pdf

LAPPI, J. Calibration of height and volume equations with random parameters. Forest Science, v. 37, n. 3, p. 781– 801, 1991.

LYNCH, T. B.; HOLLEY, A. G.; STEVENSON, D. J. A Random-Parameter Height-Dbh Model for Cherrybark Oak. *South. Southern Journal of Applied Forestry*, v. 29, n. 1, p. 22–26, 2005.

MARZILIANO, P. A.; TOGNETTI, R.; LOMBARDI, F. Is tree age or tree size reducing height increment in *Abies alba* Mill. at its southernmost distribution limit? Annals of Forest Science, v. 76, n. 17, p. 1-12, 2019.

MEHTÄTALO L.; MIGUEL, S.; GREGOIRE, T. G. Modeling height-diameter curves for prediction. Canadian Journal of Forest Research, v. 45, n. 7, p. 826-837, 2015.

MEHTÄTALO, L. lmfor: Functions for Forest Biometrics. R package version 1.6. 2022. Available at: https://cran.r-project.org/web/packages/lmfor/lmfor.pdf

NG, C. T.; LEE, W.; LEE, Y. In defense of LASSO. Communication in Statistics-Theory and Methods, v. 51, n. 9, p. 3018–3042, 2020.

OLIVEIRA, L. C.; BERTOL, I.; CAMPOS, M. L.; et al. Erosão hídrica em plantio de pinus, em estrada florestal e em campo nativo. Floresta, v. 44, n. 2, p. 239 - 248, 2014.

OLIVEIRA, X. M.; SCOLFORO, H. F.; MCTAGUE, J. P.; et al. Impact of the inclusion of the sociological position and upper-stem diameter in the taper modeling of *Pinus taeda* trees, subjected to different thinning regimes. Trees, Forests and People, v. 8, 100251, 2022.

OLIVEIRA, X. M.; SCOLFORO, H. F.; MCTAGUE, J. P.; et al. Modeling the effect of stand homogeneity on basal area growth for unthinned and thinned *Pinus taeda* L. stands in southern Brazil. New Forests, v.55, p. 1455-1474, 2024.

PINHEIRO, J. C.; BATES, D. M. Mixed-Effects Models in Sand S-PLUS. New York, NY: Springer New York. 2013. 537 p.

PINHEIRO, J.; BATES, D.; DEBROY, S.; et al. nlme: Linear and nonlinear mixed effects models. R Package version 3.1-164, 2023. Available at: https://CRAN.R-project.org/package=nlme

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2023. Available at: http://www.R-project.org

RAPTIS, D. I.; KAZANA, V.; KAZAKLIS, A.; et al. Mixed-effects height–diameter models for black pine (*Pinus nigra* Arn.) forest management. Trees, v. 35, p. 1167–1183, 2021.

RIBEIRO, A.; FERRAZ FILHO, A. C.; MELLO, J. M.; et al. Estratégias e metodologias de ajuste de modelos hipsométrircos em plantios de *Eucalyptus* sp. Cerne, v. 16, n. 1, p. 22–31, 2010.

SAUD, P.; LYNCH, T. B.; ANUP, K. C.; et al. Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data. Forestry, v. 89, n. 2, p. 215–229, 2016.

SCOLFORO, H. F.; MCTAGUE, J. P.; BURKHART, H.; et al. Eucalyptus growth and yield system: Linking individual-tree and stand-level growth models in clonal Eucalypt plantations in Brazil. Forest Ecology and Management, v. 432, p. 1-16, 2019.

SCOLFORO, J. R. S.; MAESTRI, R.; FERRAZ FILHO, A. C.; et al. Dominant height model for site classification of *Eucalyptus grandis* incorporating climatic variables. International Journal of Forestry Research, v. 10, p. 1-7, 2013.

SILVA, L. F.; FERREIRA, G. L.; SANTOS, A. C. A.; et al. Equações hipsométricas volumétricas e de crescimento para *Khaya ivorensis* plantada em Pirapora. Floresta e Ambiente, v. 23, n. 3, p. 362-368, 2016.

STRIMBU, B. M.; AMARIOAREI, A.; MCTAGUE, J. P.; et al. A posteriori bias correction of three models used for environmental reporting. Forestry, v. 91, n. 1, p. 49-62, 2018.

TIBSHIRANI, R. Regression Shrinkage and Selection via the LASSO. Journal of the Royal Statistical Society, v. 58, n. 1, p. 267–288, 1996.

VARGAS-LARRETA, B.; CASTEDO-DORADO, F.; ÁLVAREZ-GONZÁLEZ, J. G.; et al. A generalized height-diameter model with random coefficients for uneven-aged stands in El Salto, Durango (Mexico). Forestry, v. 82, n. 4, p. 445–462, 2009.

WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. R Package version 3.4.0. 2022. Available at: https://cran.r-project.org/web/packages/ggplot2/index.html