Research Article
doi: 10.1590/01047760202228013033
vol(28), 2022

CERNE

Generalized height-diameter models with random effects for natural forests of central Mexico

Ernesto Alonso Rubio Camacho [1✤iD], Sacramento Corral Rivas [2iD], Juan Antonio López Hernández [iD], Álvaro Agustín Chávez Durán [1iD], Jaqueline Xelhuantzi Carmona [1iD], Jürgen Nagel [3iD]

[1] Campo Experimental Centro Altos de Jalisco, INIFAP-CIRPAC. Tepatitlán, México.
[2] Instituto Tecnológico de El Salto. El Salto, México.
[3] Department Ecoinformatics, Biometrics & Forest Growth.Georg-August University of Goettingen. Goettingen, Germany.

FOREST MANAGEMENT

ABSTRACT

Background: Tree height is an important variable in forestry, as it is commonly used to estimate volume and biomass, and to evaluate site productivity. In this study, we developed four generalized equations to model height-diameter (*h-d*) relationships for coniferous and broadleaf species. For this purpose, we used information from 49 permanent sampling plots located in the natural forests of Puebla, Mexico. Non-linear fixed and mixed-effects modeling approaches were used to fit generalized versions of the Gompertz function to *Pinus patula* and the *Pinus* group, the Näslund function to *Abies religiosa*, and the Curtis function to the *Quercus* group.

Results: Stand variables included in the models were the number of trees per hectare (*N*), quadratic mean diameter (*dg*), and basal area per hectare (*G*). The results showed a model efficiency (EF) = 0.91 and root mean square error (RMSE) = 2.04 for *P. patula*, as well as an EF = 0.91 and RMSE = 1.63 for the *Pinus* group. The EF and RMSE for *Abies religiosa* were 0.88 and 2.21, while for the *Quercus* group these values were 0.72 and 1.9, respectively. From the mixed-effects model calibration, only a sub-sample of three trees from different quantiles of the diameter distribution is required to make accurate predictions. No stand-level variables related to tree height are included in any of the selected models, thus no additional measurements beyond tree diameter are required.

Conclusion: Compared to conventional non-linear least squares (ONLS), mixed-effects models are more flexible and accurate and represent a new tool for sustainable forest management of natural forests in the study area.

Keywords: *Pinus patula*, *Abies religiosa*, *Pinus*, *Quercus*, mixed-effects, calibration, cross-validation.

HIGHLIGHTS

Four new generalized height-diameter models were developed for species of central México.
Cross-validation was used to validate and calibrate the mixed-effects models simultaneously.
No tree height-related stand variables were used as predictors in the generalized models.
The number of trees per hectare was the most used variable and included in all selected models.

CAMACHO, E. A. R.; RIVAS, S.C.; HERNÁNDEZ, J. A. L.; DURÁN, A. A. C.; CARMONA, J. X.; NAGEL, J. Generalized height–diameter models with random effects for natural forests of central Mexico. CERNE, v.28, e-103033, doi: 10.1590/01047760202228013033

✤Corresponding author e-mail: sacra.corral@gmail.com
Submitted: 13/04/2021 Accepted: 25/11/2021

CERNE
ISSN 0104-7760
UNIVERSIDADE FEDERAL DE LAVRAS

INTRODUCTION

Tree height (*h*) is an important element in forest management and is used to characterize forest structure, estimate volume and biomass, and assess site quality (Laar and Akça, 2007; Burkhart and Tomé, 2012). However, the measurement of tree height often requires more time and effort than measuring tree diameter or other tree variables. In addition, tree height is prone to measurement errors as it requires observing the top of the tree to make precise estimates (Prodan et al., 1997), which is often complicated in closed-canopy conditions, with broken trees, and in complex topographic positions (Vanclay, 1994). Nevertheless, there are tools, such as allometric tree height-diameter (*h-d*) models, that can be used to estimate tree height instead of direct measurements (Calama and Montero, 2004; Mehtätalo et al., 2015).

Several functions, both linear and non-linear, have been used to fit allometric tree *h-d* relationships, a compilation of which can be found in Mehtätalo et al. (2015) and Ogana et al. (2020). Functions that only include tree diameter as a predictor variable are known as *local models* and can be applied to accurately predict tree height, but they are not inherently useful in all forest conditions (Laar and Akça 2007; Corral-Rivas et al., 2019). Therefore, a possible solution is to generalize *local models* by including stand variables that account for tree density or site quality (López et al. 2012; Hernández-Ramos et al., 2015). Furthermore, generalized models are often necessary in forest growth simulators (Hansen and Nagel, 2014) and for imputation of missing tree heights in large databases (Mehtätalo et al., 2015).

In general, tree height is recorded in sample plots, generating grouped data. This structure often leads to a lack of independence among observations, thus violating the assumptions of ordinary least squares (OLS) traditionally used in forestry (West el al., 1984). The mixed-effects modeling approximation accounts for grouped structures (López et al., 2012) by modeling the variability between plots (Caballero-Deloya, 1998; Castedo Dorado et al., 2006). Furthermore, mixed-effects models can be calibrated (also known as localized) for tree height prediction on new plots when at least one additional tree height and diameter (pair) measurement is available, which simplifies their use and application in forestry (Corral-Rivas et al., 2019).

In Mexico, mixed-effects models for predicting tree height have been scarcely applied, although some studies have been carried out in the state of Durango (Vargas-Larreta et al., 2009; Corral-Rivas et al., 2019). However, for central Mexico, most research studies do not include random effects (Guerra-De la Cruz et al., 2019). Therefore, deploying tree *h-d* models that account for hierarchical data structure, i.e. random effects, would contribute to Mexican forest science by providing a new management tool for economically and ecologically important species in central Mexico.

The aim of this study was to develop four generalized tree *h-d* mixed-effects models for *Pinus patula* Schiede ex Schltdl. & Cham. and other major species of pure and mixed stands in northern Puebla, Mexico. The specific objectives were: (i) to test different *local models* from the literature and select the best fitted per species or genus (four in total); (ii) to generalize the best *local models* by including stand-level predictor variables; (iii) to test different combinations of parameters to expand with random effects at plot level; and (iv) to calibrate and validate the generalized mixed-effects models.

MATERIAL AND METHODS

Study Area

This study was carried out in natural forests in northern Puebla, Mexico, located in the geographical province of the "Eje Neovolcanico Transversal". The selected forest stands are part of the Forest Management Unit 2108 (UMF2108) "Zacatlán" located at the extreme coordinates of longitude [-98.06963°; -98.49796°] and latitude [19.68858°; 19.94516°]. In this section of Puebla, the climate is mostly temperate sub-humid [12 – 18°*C*] with rains in summer (INEGI, 2008). The natural forests are mainly represented by associations of *Pinus patula*, *Pinus teocote* Schltdl. & Cham, *Pinus ayacahuite*, *P. pseudostrobus* Lindl., *Abies religiosa* (Kunth) Schltdl. & Cham., and *Quercus laurina* Humb et Bonpl.

The database used in this study comes from 49 permanent square plots (30x30 m), established in 2008 under a stratified random sampling design and remeasured in 2012 (Figure 1). In this study we used only the data from 2012. In these plots, each tree with a diameter at breast height (1.3 m) equal to or larger than 7.5 cm was measured, and data were collected on species, diameter at 1.3 m (*d*, cm), and total tree height (*h*, m), estimated with the use of a SUUNTO® clinometer. With these data, we estimated the number of trees per hectare (*N*), basal area per hectare (*G*), quadratic mean diameter (*dg*, cm), dominant tree height, as the average of the 100 thickest trees per hectare (H_d, m), and its corresponding dominant tree diameter (D_d, cm) (Assmann 1970).

The species composition in the database was as follows: 81% *Pinus* (*P. ayacahuite* Ehrenb., *P. montezumae* Lamb., *P. patula*, *P. pseudostrobus*, and *P. teocote*), 7% *Quercus* (*Q. rugosa* Née and *Q. laurina*) and 11% *Abies religiosa*. Table 1 summarizes the descriptive statistics of the database. As some species have few observations, the individuals were grouped and fitted as follows: (i) *P. patula*, (ii) *Pinus* group (*P. ayacahuite*, *P. montezumae*, *P. pseudostrobus*, and *P. teocote*), (iii) *Abies religiosa*, and (iv) *Quercus* group (*Q. rugosa* and *Q. laurina*).

Model Development

The tree *h-d* modeling process was divided into three steps: the first one was the *d*evelopment of the generalized mixed-effects models, where we fitted the six *local models* from the literature (Table 2, models 1-6) and selected one per group. Then, the selected models were generalized by directly adding stand variables from Table 1 into the model. The second step was the fitting (using the mixed modeling approximation) of the generalized models from the literature (Table 2, models 7-12) to compare them with the model developed in the first step. The third step comprised the calibration and validation of selected mixed-effects generalized models, one per group, using the cross-validation. These processes are described in more detail below.

Tree *height-diameter* functions

To select the models that best explain the *h-d* relationship per group (*P. patula*, *A. religiosa*, *Pinus*, and *Quercus* groups), we tested 12 non-linear functions frequently used in forestry (López et al. 2012; Corral-Rivas et al. 2014; Hernández-Ramos et al. 2015; Ogana et al. 2020) (Table 2), using the mixed-effects modeling approximation.

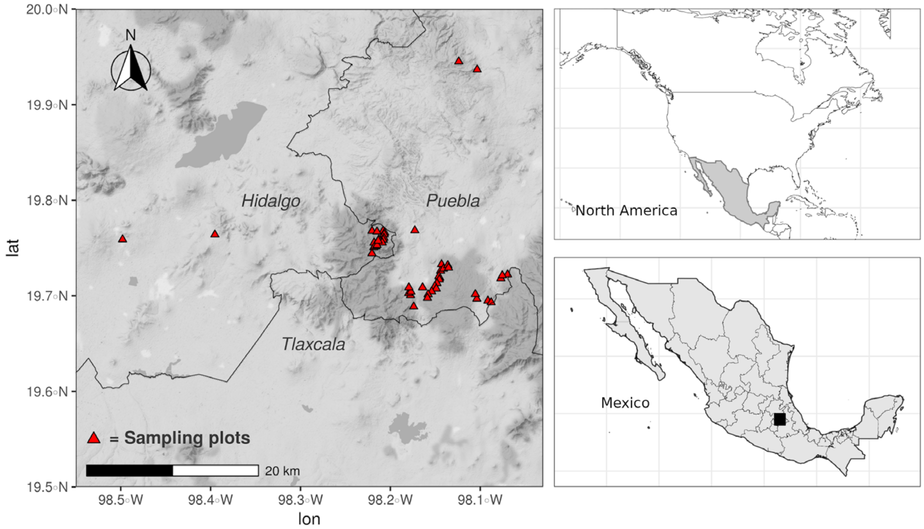

Figure 1. Study area location.

Table 1. Descriptive statistics of the database by group and stand. Where *d* = mean tree diameter in cm, *h* = mean tree height in m, *N* = number of trees per hectare, *G* = basal area per hectare in m^2, *dg* = quadratic mean diameter in cm, H_o = dominant tree height in m, D_o = dominant tree diameter in cm, SD = standard deviation.

	Obs.	min	max	mean	SD
Pinus patula					
d (cm)	1594	7.5	78.5	26.6	13.6
h (m)		5.44	39	19.9	6.7
Pinus					
d (cm)	625	7.5	67.6	17.6	8.3
h (m)		4.75	33.5	11.8	5.5
Abies religiosa					
d (cm)	302	7.5	81.6	23.6	13.4
h (m)		5.25	40.25	18.9	6.3
Quercus					
d (cm)	204	7.5	33.4	15	4.8
h (m)		2.9	22.5	9.9	3.6
Stand					
N	2725	200	2311	981	571
G		10.9	84.3	43.7	13.8
dg		13.7	52.4	26.3	7.6
Ho		8.2	38.3	24.4	7.5
Do		16.9	66.7	36.5	13.3

Table 2. The tree *h-d* equations analyzed in this study. Where h = total height (m); d = tree diameter (cm); H_0 = dominant tree height (m); D_0 = dominant tree diameter (cm); N = number of trees per hectare; G = basal area per hectare (m^2. ha^{-1}); dg = quadratic mean diameter (cm); hg = height corresponding to quadratic mean diameter (m); β_i = model parameters.

Source	Expression	No.
Curtis, (1967)	$h = 1.3 + \frac{\beta_0 \cdot d}{(1+d)^{\beta_1}}$	(1)
Schumacher, (1939)	$h = 1.3 + \beta_0 \cdot exp(-\beta_1 \cdot d^{-1})$	(2)
Näslund, (1936)	$h = 1.3 + \frac{d^2}{(\beta_0 + \beta_1 \cdot d)^2}$	(3)
Richards, (1959)	$h = 1.3 + \beta_0 \cdot \left(1 - exp(-\beta_1 \cdot d)\right)^{\beta_2}$	(4)
Gompertz, (1825)	$h = 1.3 + \beta_0 \cdot exp\left(-\beta_1 \cdot exp(-\beta_2 \cdot d)\right)$	(5)
Weibull (Weibull, 1951)	$h = 1.3 + \beta_0 \cdot \left(1 - exp\left(-\beta_1 \cdot d^{\beta_2}\right)\right)$	(6)
Nagel (Nagel, 1999)	$h = 1.3 + (hg - 1.3) \cdot exp\left(\beta_0 \cdot \left(1.0 - \left(\frac{dg}{d}\right)\right)\right) \cdot exp\left(\beta_1 \cdot \left(\left(\frac{1}{dg}\right) - \left(\frac{1}{d}\right)\right)\right)$	(7)
Schröder-Álvarez I (2001)	$h = 1.3 + (\beta_0 + \beta_1 \cdot H_0 - \beta_2 \cdot dg) \cdot exp\left(\frac{-\beta_3}{\sqrt{d}}\right)$	(8)
Schröder-Álvarez II (2001)	$h = 1.3 + (\beta_0 + \beta_1 \cdot H_0 - \beta_2 \cdot dg + \beta_3 \cdot G) \cdot exp\left(\frac{-\beta_4}{\sqrt{d}}\right)$	(9)
Ogana et al., (2020)	$h = 1.3 + \left(\frac{\frac{d^{\beta_2}}{exp(-\beta_4)}}{log(H_0)} \cdot \frac{\beta_1 + d^3}{\beta_2}\right) \cdot exp\left(\frac{-\beta_3}{\sqrt{\left(\frac{N}{dg}\right)}}\right)$	(10)
Corral et al., (2019)	$h = 1.3 + \beta_0 \cdot log(H_0)^{\beta_1} \cdot \left(1 - exp(-\beta_2 \cdot d)\right)^{\beta_3} \cdot \left(\frac{N}{G}\right)^{\beta_4}$	(11)
Sharma et al., (2016)	$h = 1.3 + \beta_0 \cdot {H_0}^{\beta_1} \cdot \left(1 - exp\left(-\beta_2 \cdot dg^{\beta_3} \cdot d\right)\right)^{\beta_4}$	(12)

Mixed-effects modeling approach

Mixed-effects models incorporate variables as fixed and random. In tree *h-d* modeling, while the population response for the fixed-effects model (when random effects are zero) are common to all trees and plots, random effects are specific to each plot, or to any other predefined grouping factor (Pinheiro and Bates 2000; Mehtätalo et al. 2015). In this study, we used the mixed-effects modeling approach to account for tree height variation within and between plots, incorporating the random effect at plot level. We also used this approach to overcome the lack of independence between observations (Castedo-Dorado et al. 2006; Corral-Rivas et al. 2019). According to Pinheiro and Bates (2000), a mixed-effects model can be defined as follows:

$$y_{ij} = f\left(\beta_i, x_{ij}\right) + e_{ij} \quad (13)$$

Where: y_{ij} = the *j-th* height (tree) taken at the *i-th* plot, x_{ij} = the *j-th* measurement of the predictor variable taken from the *i-th* plot, β_i = parameter vector $r \times 1$ (r is the total number of parameters in the model) specific to each plot, f = non-linear function, e_{ij}= error term.

The parameter vector β_i can be divided into the fixed part, common to the whole population, and the random components, specific to each sampling plot with the form: $\beta_i = A_i\lambda + B_i b_i$ where λ is the vector $p \times 1$ of the fixed parameters, (p is the number of fixed parameters); b_i is the vector $q \times 1$ of the random parameters associated with the *i-th* plot (q is the number of random parameters), A and B are design matrices of size $r \times p$ and $r \times q$ for the fixed and random effects specific to the *i-th* plot, respectively. The basic theoretical assumptions of non-linear mixed-effects models assume that the vector of residuals (e_i) and the vector of the random effect (b_i) have a normal distribution with mean equal to zero and a variance-covariance matrix (R_i and D) representing the variability among the different sampling plots (Lindstrom and Bates 1990; Littell et al. 2006).

To carry out the parameter estimation, the non-linear models were linearized using the first-order Taylor expansion around the random effects (Pinheiro and Bates 2000). The Lindstrom-Bates algorithm was used in the fitting process through the first-order conditional expectation (FOCE) approximation as the expansion method (Lindstrom and Bates 1990).

Parameter expansion with random effects

The procedure we followed, to choose which parameters to expand with random effects, was to consider all parameters from the non-linear model as fixed and random (mixed). When the model failed to converge, we systematically took out random effects until convergence was reached (Pinheiro and Bates 2000). In our case, several combinations of random effects were tested and the final model was chosen based on the goodness-of-fit statistics.

In the fitting process, we chose the restricted maximum likelihood method (REML) to obtain the model parameters with the *nlme* package implemented in R (R Core Team 2019).

Calibration of the mixed-effects model.

In forestry, the application of an already fitted mixed-effects model to a new plot is usually done by a method called calibration or localization (Hall and Bailey 2001). In this study, the calibration involved the use of at least one tree height measurement from the new plot to predict the random effects. We used the Lindstrom-Bates algorithm through the FOCE method in the calibration procedure, where the value of the vector of random parameters b_j, associated with the sampling plot, can be estimated through the following equation (Vonesh and Chinchilli 1997):

$$\hat{b} = \hat{D}\hat{Z}_i^T\left(\hat{R}_i + \hat{Z}_i\hat{D}\hat{Z}_i^T\right)^{-1}\hat{e}_i \quad (14)$$

Where: $\hat{D}$= variance-covariance matrix $q \times q$ associated with the random parameters, (q is the number of random parameters included in the model) common to all plots. $\hat{R}_i$ = variance-covariance matrix $m_j \times m_j$ of the model error term, $\hat{e}_i$ = vector of residuals $m \times 1$, whose components are obtained by the difference between the observed height of each tree and the predicted value using the model with only fixed parameters. $\hat{Z}_i$ = matrix $m \times q$ of the partial derivatives of the random parameters evaluated in $\hat{b}$=0.

Number of trees for calibration

In order to evaluate the number of trees in the calibration process, we followed a similar approach to Ogana et al. (2020) and Corral-Rivas et al. (2019), i.e. we tested different subsample combinations based on the diameter distribution. In this regard, we selected different sample trees per plot under the following combinations: three sample trees $Q_{[1;2;3]}$ = taking a subsample of the three closest trees to the 0.25, 0.50, and 0.75 quantiles of the diameter distribution. The remaining subsample categories include the $Q_{[1;2]}$, $Q_{[1;3]}$, $Q_{[2;3]}$, $Q_{[2]}$, and the three trees closest to the diameter d_g, the smallest [min] and the largest [max], respectively.

Calibration was evaluated by estimating the amount of the root mean squared error (RMSE) reduction given by the different calibration options. Then, we compared the calibration options with the *Nlme* when all trees in the sampling plot are considered to estimate the random parameters. Using this approach, we were able to evaluate the number of trees using different calibration options; this process could also be combined with the model validation step.

Validation of the mixed-effects model

In this study, we used leave-one-out cross-validation (Hastie et al. 2009) and calibration simultaneously to evaluate the models. For this purpose, we coded a *for loop* in R that follows the next steps: step 1) take out the plot number one and fit the mixed-effects model using the remaining plots; step 2) take a subsample from plot one and estimate the random effects using equation (13); step 3) use the predicted random effects along with the fixed effects from step 1 and make tree height predictions on the remaining trees; step 5) repeat the process with the plot number two and continue with the remaining plots until the whole database has all tree heights predicted; and step 6) with the observed and predicted tree heights calculate the RMSE. Overall, it was possible to use the results to rank the model based on the cross-validated/calibrated RMSE.

Model assessment

The best models were selected on the basis of model accuracy and parsimony. While the former was evaluated using the RMSE and model efficiency (EF), which is similar to the coefficient of determination R^2, parsimony was evaluated through the Akaike and Schwarz Bayesian information criteria, *AIC* and *BIC*, respectively (Schwarz 1978), which are expressed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}} \quad (15)$$

$$AIC = n\,log\left(\frac{SSR}{n}\right) + 2p \quad (16)$$

$$EF = 1 - \frac{(n-1)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-p)\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (17)$$

$$BIC = n\,log\left(\frac{SSR}{n}\right) + p\,log(n) \quad (18)$$

Where: y_i = observed values, y_i= predicted values, $\hat{y}_i$ = average, n = total number of observations, p = number of parameters of the equation, *SSR* = sum of the squares of the residuals, *log* = natural logarithm.

RESULTS

Overall, given the results we selected the proposed generalized mixed models instead of those from the literature. On the one hand, the proposed models have lower AIC and BIC, although the difference is not that great. On the other hand, the selected models do not include tree height-related stand variables, which can reduce the workload and make these models easier to apply, especially in the Mexican context. Further details of the results are given in the following paragraphs.

Local mixed-effects models

The six *local models* (models 1 to 6 in Table 2) had different degrees of accuracy depending on the group and the number of parameters. The Gompertz model (5) was selected for the *P. patula* and *Pinus* groups, whereas for the *A. religiosa* and *Quercus* groups the best models were the Näslund (3) and Curtis (1) ones, respectively. The decision was based on the AIC and BIC (Table 3) ranking and visual analysis of the graphs (Figure 2).

Table 3. Fitting results of the selected local mixed-effects models per group. Where β_0..β_2 are the fixed-effects parameters; u_i and v_i are the random effects.

Parameters	*Pinus patula*		*Pinus group*		*Abies religiosa*		*Quercus group*	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	29.06	0.957	23.04	1.63	1.94	0.14	22.64	1.70
β_1	2.01	0.131	2.30	0.22	0.15	0.01	15.05	0.98
β_2	0.06	0.002	0.08	0.01	-	-	-	-
$sd(u_i)$	5.97	-	6.47	-	0.55	-	3.98	-
$sd(v_i)$	0.75	-	0.81	-	0.02	-	0.0004	-
$Cor(u_i, v_i)$	0.81	-	0.81	-	-0.97	-	0.0020	-
AIC	6791.50	-	2384.50	-	1336.65	-	839.60	-
BIC	6812.99	-	2402.25	-	1347.78	-	849.55	-

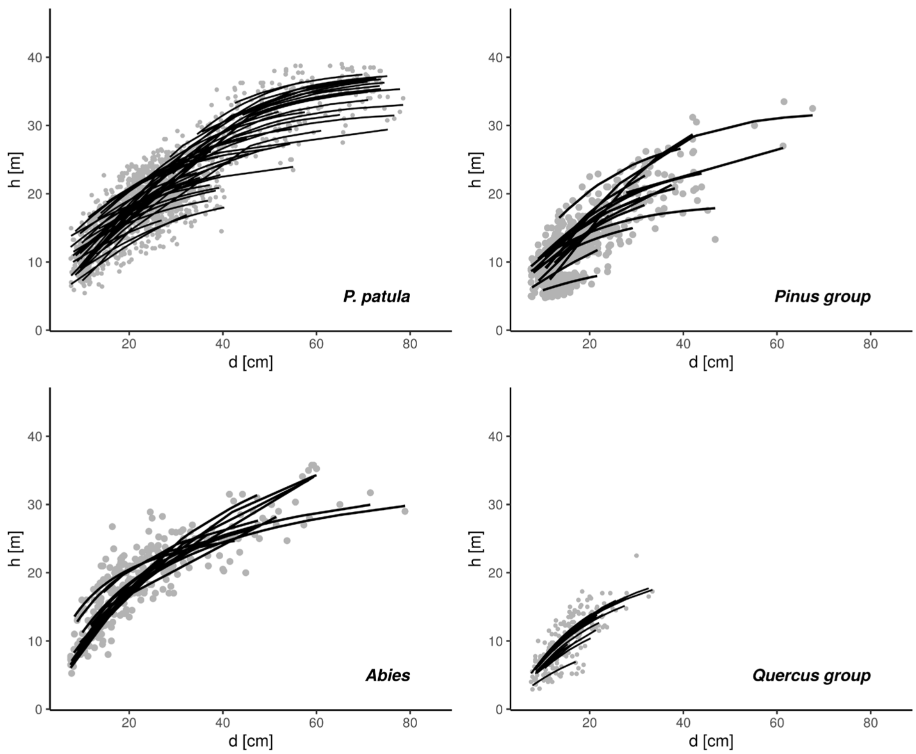

Figure 2. Graphical fitting of the selected local mixed-effects models. Where the black lines represent the local curves and the dots are the observed h-d data.

Generalized mixed-effects models

Pinus patula. For *P. patula*, the Gompertz model (5) was generalized by including quadratic mean diameter (log(dg)), basal area per hectare (G), and number of trees per hectare (N). log(d_g) was incorporated to β_0 and the log(G/N) ratio to the β_1 parameter. Although the inclusion of mean and dominant heights generally improves the height-diameter models, a large number of trees are needed to estimate such variables, which implies additional sampling cost. Regarding random effects, although we tested different combinations, only two parameters could be expanded with random effects β_0 and β_1. Therefore, the generalized mixed-effects model for *P. patula* was fitted through the following expression:

$$h_{ij} = 1.3 + (\beta_0 + u_i + \beta_3 \cdot \log(d_g))\, exp\left(-\left(\beta_1 + v_i + \beta_4 \cdot log\left(\frac{G_i}{N_i}\right)\right) \cdot exp(-\beta_2 \cdot d_{ij})\right) + e_{ij} \quad (19)$$

Where h_{ij} is the tree height of the j tree in the i sample plot, d_{ij} is the tree diameter of the j tree in the i sample plot, dg_i is the quadratic mean diameter for plot i, G_i and N_i are the basal area per hectare and the number of trees per hectare for plot i, $\beta_0 \ldots \beta_4$ are the fixed-effects parameters, u_i and v_i are random effects, and e_{ij} is the model error.

The proposed generalized mixed-effects model (19) was compared with the generalized ones from the literature (models 7 to 12 from Table 2). In general, although the models had only marginal differences, we selected model (19) as the final model because it has the lowest values of RMSE, AIC, and BIC (Table. 4). In addition, model (19) does not include stand-level predictors associated with tree height, i.e. in practice, it does not need additional tree height measurements other than those used in the calibration procedure. Figure 3 shows no systematic trend in the distribution of the residuals.

Table 4. Fitting results of the selected generalized mixed-effects model for *Pinus patula*. Where $\beta_0 \ldots \beta_4$ are the fixed-effects parameters; u_i and v_i are the random effects.

Pinus patula (Model 19)			
Parameters	Estimate	SE	t-value
β_0	-35.41	7.97	-4.45
β_1	4.68	0.55	8.52
β_2	0.06	0	23.12
β_3	18.87	2.33	8.11
β_4	1.01	0.19	5.36
Random components			
$sd(u_i)$	3.72	-	-
$sd(v_i)$	0.47	-	-
$Cor(u_i,v_i)$	0.45	-	-
Goodness-of-fit statistics			
EF	0.91	-	-
RMSE	2.04	-	-
AIC	6800.74	-	-
BIC	6832.99	-	-

Pinus group

The generalized model for the *Pinus* group, based on the Gompertz function (5), includes the quadratic mean diameter (d_g) and the number of trees per hectare (log(N)) incorporated to β_0 and to β_1, respectively. In relation to random effects, as well as for *P. patula*, only two parameters could be expanded with random effects, β_0 and β_1. Thus, the generalized mixed-effects model for the *Pinus* group was fitted through the following expression:

$$h_{ij} = 1.3 + (\beta_0 + u_i + \beta_3 \cdot dg_i \;\; exp\left(-(\beta_1 + v_i + \beta_4 \cdot log(N_i)) \cdot exp(-\beta_2 \cdot d_{ij})\right) + e_{ij} \quad (20)$$

The generalized mixed-effects model for the *Pinus* group (20) was selected based on the goodness-of-fit statistics (Table. 5). Model (20) converged by expanding two parameters with random effects. Although the generalized models have only marginal differences with other models, we selected model (20) as the final model because it has the lowest values of RMSE, AIC, and BIC and does not include stand-level predictors associated with tree height. As in the case of *P. patula*, the regression line for the *Pinus* group shows no evidence of heteroscedasticity in the distribution of the residuals (Figure 3).

Abies religiosa

The generalized model for *Abies religiosa* was based on the Näslund model (3). Stand-level predictors included the log(G/N) ratio associated with β_0 and the quadratic

Table 5. Fitting results of the selected generalized mixed-effects model for the *Pinus* group. Where $\beta_0 \ldots \beta_4$ are the fixed-effects parameters; u_i and v_i are the random effects.

Pinus Group (Model 20)			
Parameters	Estimate	SE	t-value
β_0	12.69	3.30	3.84
β_1	7.74	1.61	4.81
β_2	0.08	0.01	12.11
β_3	0.34	0.11	3.13
β_4	0.83	0.23	3.56
Random components			
$sd(u_i)$	4.92	-	-
$sd(v_i)$	0.52	-	-
$Cor(u_i,v_i)$	0.64	-	-
Goodness-of-fit statistics			
EF	0.91	-	-
RMSE	1.63	-	-
AIC	2392.60	-	-
BIC	2419.23	-	-

mean diameter (d_g) with β_1. Regarding random effects, similar to the previous groups, only two parameters could be expanded with random effects, β_0 and β_1. Therefore, the generalized mixed-effects model for *A. religiosa* was fitted by using the following expression.

$$h_{ij} = 1.3 + \frac{d_{ij}^2}{\left((\beta_0 + u_i + \log\left(\frac{G_i}{N_i}\right) \cdot \beta_2) + (\beta_1 + v_i + \log(dg_i) \cdot \beta_3) \cdot d_{ij}\right)^2} + e_{ij} \quad (21)$$

Compared to the generalized models from the literature, model (21) was ranked as the best one and chosen as the final model (Table 6). Additionally, the regression results for *A. religiosa* do not show any systematic trend in the distribution of the residuals (Figure 3).

Table 6. Fitting results of the selected generalized mixed-effects model for *Abies religiosa*. Where $\beta_0 \ldots \beta_4$ are the fixed-effects parameters; u_i and v_i are the random effects.

Abies religiosa (Model 21)			
Parameters	Estimate	SE	t-value
β_0	3.58	0.74	4.85
β_1	0.32	0.07	4.38
β_2	0.65	0.28	2.31
β_3	-0.05	0.02	-2.35
Random components			
$sd(u_i)$	0.44	-	-
$sd(v_i)$	0.02	-	-
$Cor(u_i,v_i)$	-0.94	-	-
Goodness-of-fit statistics			
EF	0.88	-	-
RMSE	2.21	-	-
AIC	1342.85	-	-
BIC	1361.40	-	-

Quercus group

The generalized model proposed for the *Quercus* group was based on the Curtis model (1) and only included the square root ratio of *N* and *G* as stand-level predictors. Two parameters were expanded with random effects, β_0 and β_1. Therefore, the generalized mixed-effects model for the *Quercus* group was fitted through the following expression:

$$h_{ij} = 1.3 + \frac{\left(\beta_0 + u_i + \sqrt{\frac{N_i}{G_i}} \cdot \beta_2\right) \cdot d_{ij}}{(1 + d_{ij})^{\beta_1 + v_i}} + e_{ij} \quad (22)$$

The generalized mixed-effects model for the *Quercus* group (22) showed the best goodness-of-fit statistics compared to the generalized models from the literature (Table 7). Therefore, this model was chosen as the final one.

Table 7. Fitting results of the selected generalized mixed-effects model for the *Quercus* group. Where $\beta_0 \ldots \beta_4$ are the fixed-effects parameters; u_i and v_i are the random effects.

Quercus group (Model 22)			
Parameters	Estimate	SE	t-value
β_0	31.53	3.79	8.31
β_1	15.08	0.98	15.35
β_2	-2.48	0.95	-2.61
Random components			
$sd(u_i)$	3.00	-	-
$sd(v_i)$	0.0002	-	-
$Cor(u_i,v_i)$	0.0010	-	-
Goodness-of-fit statistics			
EF	0.72	-	-
RMSE	1.90	-	-
AIC	846.68	-	-
BIC	859.96	-	-

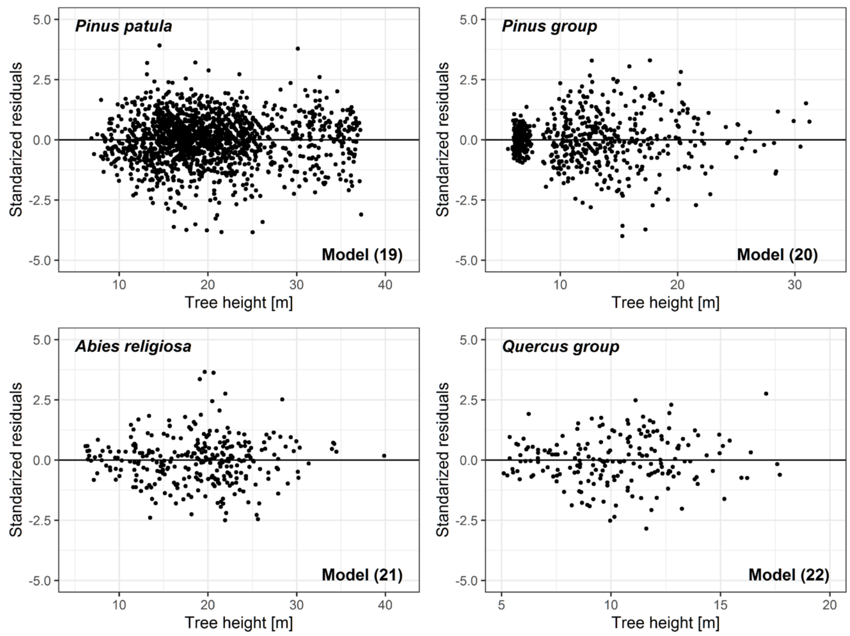

Figure 3. Residual plots per group of the generalized mixed-effects models.

Calibration and validation

Overall, the results regarding the evaluation of the calibration approach showed that in all cases it is better to use three trees to predict random effects. For *P. patula* and the *Pinus* group, the best option was achieved by measuring the total tree height of the three trees: 1) the tree with diameter closer to the *dg*, 2) the smaller (min *d*), and 3) the larger tree (max *d*) (Table 8). The best calibration strategy for *Abies religiosa* and the *Quercus* group was achieved by measuring the total height of three trees with a diameter close to the first three quantiles of the diameter distribution (Table 8).

DISCUSSION

Local fixed-effects *h-d* models

Among the six local functions evaluated (Table 2), the Gompertz function (Gompertz, 1825) was the most accurate for *P. patula* and the *Pinus* group. The Gompertz function has a wide range of application in forest growth modeling (Burkhart and Tomé, 2012; Pödör et al. 2014). It has also been previously applied to describe the tree *h-d* relationships in natural forests in Nigeria (Ogana, 2019) and to conifer species in the inland Northwest of the United States (Zhang, 1997), as well as to analyze how the tree *h-d* allometry is related to climate in the United States (Hulshof et al. 2015). In our study, considering the results on goodness-of-fit and model parsimony, the Gompertz function (5) was chosen for further analysis for *P. patula* and the *Pinus* group.

Näslund (Näslund, 1936) (3) and Curtis (Curtis, 1967) (1) functions were the most suitable for *A. religiosa* and the *Quercus* group, respectively. On the one hand, the Näslund function has been widely applied to describe tree *h-d* relationships; for example, in the studies of Mehtätalo et al. (2015) and Sharma et al. (2016), it was one of the most appropriate functions for modeling the tree *h-d* relationship in several datasets. On the other hand, the Curtis function has been tested in several studies (Mehtätalo et al., 2015; Sharma et al., 2016; Ogana, 2019). However, contrary to our case, the Curtis function usually tended to show lower goodness-of-fit compared to other tree *h-d* functions (Sharma et al., 2016; Ogana 2019), particularly when compared to 3-parameter models (Ng'andwe et al., 2019). In this study, we generalized the Näslund and Curtis functions using stand-level variables for *A. religiosa* and the *Quercus* group, respectively.

Tree *h-d* model generalization

The generalization of tree *h-d* models is necessary to account for stand density, competition, and site quality (Prodan et al. 1997; Corral-Rivas et al. 2019; Bronisz and Mehtätalo, 2020). For this purpose, stand-level variables such as quadratic mean diameter (*dg*) or dominant tree height (H_0) must be included in *h-d* models. For example, a newly developed model for *Pinus nigra*, based on the Chapman-Richards function, included dominant tree height and dominant tree diameter as stand predictor variables (Raptis et al., 2021). Corral-Rivas et al. (2019) included H_0, *G*, and *N*, while Ogana et al. (2020), Sharma et al. (2016), and Schröder and Álvarez (2001) used H_0 and *dg* as predictor variables.

In this study, we did not include stand-level variables related to tree height, such as dominant tree height. The main reason was that the integration of those stand-level predictors would imply additional field work, especially in the Mexican context. In Mexico, while most forest inventories require all tree diameters to be measured, not all of them require all tree heights. Therefore, the use of the models presented here are suitable.

In this study, the stand-level predictors included in all final generalized models can be estimated through tree diameters and the number of trees. A common variable in all four generalized models was the number of trees per hectare (*N*), while *dg* was added to *P. patula*, *A. religiosa*, and *Pinus* models, and *G* was included in *P. patula*, *A. religiosa*, and *Quercus* models. Similar studies included *dg* and *G* as stand-level predictor variables as these did not require additional measurements except tree diameter (Mehtätalo et al. 2015; Bronisz and Mehtätalo 2020).

Parameter expansion with random effects

The parameter expansion procedure included the exploration of different parameter combinations. As recommended by Sharma and Parton (2007) and Mehtätalo et al. (2015), the first step was to try to fit

Table 8. Comparison of the calibration alternatives applied to the selected generalized mixed-effects models. The table shows the RMSE using different combinations of sub-sample tree heights. The presented values were estimated via cross-validation.

Subsample	*Pinus patula* Model (19)	*Pinus group* Model (20)	*Abies religiosa* Model (21)	*Quercus group* Model (22)
Q[1;2;3]	2.54	2.16	2.89	2.10
Q[1;2]	2.64	2.53	2.90	2.24
Q[1;3]	2.79	2.18	2.91	2.13
Q[2;3]	2.47	2.39	3.35	2.13
Q[2]	2.58	2.87	3.26	2.28
dg, dmin, dmax	2.41	2.16	3.22	2.15

the mixed-effects model using the three parameters expanded with random effects. However, convergence was not reached in any case, so it was decided to reduce the number and systematically combine parameters until convergence was achieved. Other studies have pointed out convergence problems, especially in models with more than two parameters to expand (Mehtätalo et al., 2015; Corral-Rivas et al., 2019; Ogana et al., 2020). In our study, in all the proposed generalized models, convergence was achieved with the expansion of two parameters.

Random effects calibration

In this research, we used calibration and cross-validation simultaneously to evaluate the prediction of random effects on new plots. For this purpose, it was necessary to take a sub-sample of tree heights from the new plot using an iterative approach to evaluate different calibration options (from one to three trees). In this context, some studies examined different calibration alternatives and although the results share some similarities, they also have some differences.

In the study of Corral-Rivas et al. (2019), different sub-sample options were analyzed, from one to three trees. The best calibration option was to randomly select the three closest trees to the second quantile of the diameter distribution. In the study of Ogana et al. (2020), the best strategy was to randomly select just one tree. Calama and Montero (2004) used a sub-sample of four trees. In the study of Bronisz and Mehtätalo (2020), they tested several combinations and found that the best strategy was to select the thinnest and thickest trees to predict random effects. This finding was similar to that of our study, where the best strategy was to use the thinnest, the thickest, and the closest to the quadratic mean diameter for *P. patula* and the *Pinus* group. Although many studies have shown that only a small sub-sample of trees is necessary to make accurate predictions, the number and size of trees in the calibration is not consistent and may vary depending on the final mixed-effects model. However, most studies agree that the use of mixed-effects models is justified when at least one additional tree height measurement is available.

It is important to note that all proposed models require knowledge of the diameter distribution to be calibrated correctly. Therefore, it may involve additional field work in certain cases. However, in the Mexican context, in most situations, all tree diameters are measured. In addition, the models presented in this study are intended to be integrated into dedicated software for forest management planning, and therefore do not involve additional field work in these situations.

New approaches in tree *h-d* modeling

In the case of multi-species and multi-layer forests there are cases where a dummy variable has been used instead several models (Sharma et al., 2018; Sharma et al., 2016). This may have some advantages, i) a common single model would be applicable to all species, ii) it allows to include species with only few observations, and iii) it could facilitate field work.

In our case having a single model does not necessarily have advantages over the presented generalized models. On the one hand, the development of the different specific models shown in the study has made it possible to highlight the differences in the growth patterns of the species, differences that affect the *h-d* relationship and that are manifested in the fact that the base equations and the generalizations obtained are different in each case. From a practical point of view, the use of a single equation also does not bring any improvement, since it is still necessary to distinguish between species in order to assign values of 0 or 1 to the dummy variables. Therefore, further research is needed to test this approach, especially in the Mexican forestry context.

CONCLUSIONS

In this work, we have developed four generalized mixed-effects models that describe the tree height-diameter relationship for *Pinus patula*, *Abies religiosa*, and the *Pinus* and *Quercus* groups. In general, the proposed models showed appropriate goodness-of-fit statistics and with no evident violations of the statistical assumptions. These generalized models do not use tree height related stand-level variables as predictors, i.e. except for the calibration trees, no extra height measurements of the new plot are needed to make accurate predictions.

Regarding model calibration, the best option is to estimate the random effects by only measuring the total height of three trees in the new plot. We recommend these models for total height prediction of trees growing in natural forests of northern Puebla, Mexico, specifically in the Forest Management Unit (UMAFOR) 2108, Zacatlán. However, their use can be extended to other regions where *Pinus patula* is naturally distributed, in which case we encourage users to first test the models by calibrating them.

We are confident that our research will serve as a basis for future studies on other species and other regions. We suggest that further research should be undertaken at the national forest inventory level because the use of mixed-effects models for predicting tree heights or other variables (e.g. crown width, crown base height, and crown length) will reduce cost and effort by estimating heights instead of directly measuring them. In addition, it will be possible to use these models for imputation of missing tree heights in large databases, such as those from the national forest inventory of Mexico. Finally, the generalized mixed-effects models developed in this study will contribute to Mexican forest science by providing new tools that support sustainable forest management, specifically by reducing the cost and effort involved in undertaking forest inventories.

ACKNOWLEDGEMENTS

The authors are grateful to CONAFOR Puebla for providing the data. The first author is especially thankful to CONACYT and DAAD for the joint scholarship (91680941) awarded during his doctoral studies. We thank Ulises Diéguez-Aranda and Manuel Arias-Rodil for their contribution to the R-scripts.

AUTHORSHIP CONTRIBUTION

Project Idea: EARC
Funding: EARC
Database: EARC, JALH
Processing: EARC, SCR
Analysis: EARC, SCR
Writing: EARC, SCR
Review: JALH, AACD, JXC, JN

REFERENCES

ASSMANN, E. The principles of forest yield study: studies in the organic production, structure, increment and yield of forest stands. New York, NY, USA: Pergamon Press, 1970. 506 p.

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, n. 117901, p. 1–9, 2020.

BURKHART, H. E.; TOMÉ, M. Modeling Forest Trees and Stands. Dordrecht: Springer Netherlands, 2012. 457 p.

CABALLERO-DELOYA, M. El Inventario Forestal de Mexico: Evolución y Perspectivas. Guadalajara, Jalisco, Mexico: USDA-FOREST SERVICE, 1998. Available at: http://www.fs.fed.us/rm/pubs/rmrs_p012/rmrs_p012_186_189.pdf

CALAMA, R.; MONTERO, G. Interregional nonlinear height–diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research, v. 34, n. 1, p. 150–163, 2004.

CASTEDO DORADO, F. 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.

CORRAL-RIVAS, S. et al. Local and generalized height-diameter models with random parameters for mixed, uneven-aged forests in Northwestern Durango, Mexico. Forest Ecosystems, v. 1, n. 1, p. 6, 2014.

CORRAL-RIVAS, S.; SILVA-ANTUNA, A. M.; QUIÑONEZ-BARRAZA, G. A generalized nonlinear height–diameter model with mixed-effects for seven *Pinus* species in Durango, Mexico. Revista Mexicana de Ciencias Forestales, v. 10, n. 53, p. 86–117, 2019.

CURTIS, R. O. Height-Diameter and Height-Diameter-Age Equations For Second-Growth Douglas-Fir. Forest Science, v. 13, n. 4, p. 365–375, 1967.

GOMPERTZ, B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Philosophical Transactions of the Royal Society of London, v. 115, p. 513–583, 1825.

GUERRA-DE LA CRUZ, V. et al. Modelos locales altura-diámetro para *Pinus* montezumae Lamb. y *Pinus teocote* Schiede ex Schltdl. en Nanacamilpa, Tlaxcala. Revista Mexicana de Ciencias Forestales, v. 10, n. 51, p. 133–156, 2019.

HALL, D. B.; BAILEY, R. L. Modeling and Prediction of Forest Growth Variables Based on Multilevel Nonlinear Mixed Models. Forest Science, v. 47, n. 3, p. 311–321, 2001.

HANSEN, J.; NAGEL, J. Waldwachstumskundliche Softwaresysteme auf Basis von TreeGrOSS - Anwendung und theoretische Grundlagen. Göttingen, Germany: Universitätsverlag Göttingen, 2014.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. Second ed. New York, NY: Springer New York, 2009. 745 p.

HERNÁNDEZ-RAMOS, J. et al. Ecuaciones altura-diámetro generalizadas para *Pinus teocote* Schlecht. & Cham. en el estado Hidalgo. Rev. Mex. Cienc. Forestales, v. 6, n. 31, p. 8–21, 2015.

HULSHOF, C. M.; SWENSON, N. G.; WEISER, M. D. Tree height-diameter allometry across the United States. Ecology and Evolution, v. 5, n. 6, p. 1193–1204, 2015.

INEGI. Conjunto de datos vectoriales escala 1:1 000 000. Unidades climáticas. Aguascalientes. Available at: https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825267568

LAAR, A. VAN; AKÇA, A. Forest Mensuration. Dordrecht: Springer Netherlands, 2007. v. 13. 383 p.

LINDSTROM, M. J.; BATES, D. M. Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics, v. 46, n. 3, p. 673, 1990.

LITTELL, R. C. et al. SAS for mixed models. 2nd ed.: SAS Institute. Cary, NC, USA: SAS Institute, 2006. 814 p.

LÓPEZ, S. C. A.; RODRÍGUEZ, R. S.; ÁLVAREZ-GONZÁLEZ, J. G. Relación altura-diámetro con parámetros aleatorios para rodales regulares de Pseudotsuga menziesii en el norte de España. Cuadernos de la Sociedad Española de Ciencias Forestales, n. 34, p. 135–140, 2012.

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

NAGEL, J. Konzeptionelle Überlegungen zum schrittweisen Aufbau eines waldwachstumskundlichen Simulationssystems für Nordwestdeutschland. Göttingen: Sauerländer, 1999. 122 p.

NÄSLUND, M. Skogsförsöksanstaltens gallringsförsök i tallskogMeddelanden från Statens skogsförsöksanstalt (0283-3093). Stockholm. 169 p.

NG'ANDWE, P. et al. Modeling the height-diameter relationship of planted *Pinus* kesiya in Zambia. Forest Ecology and Management, v. 447, p. 1–11, 2019.

OGANA, F. N. Tree height prediction models for two forest reserves in Nigeria using mixed-effects approach. Tropical Plant Research, v. 6, n. 1, p. 119–128, 2019.

OGANA, F. N.; CORRAL-RIVAS, S.; GORGOSO-VARELA, JOSE JAVIER. Nonlinear mixed-effect height-diameter model for *Pinus* pinaster ait. and *Pinus* radiata d. Don. CERNE, v. 26, n. 1, p. 150–161, 2020.

PINHEIRO, J. C.; BATES, D. M. Mixed-effects models in S and S-PLUS. New York: Springer-Verlag, 2000. 528 p.

PÖDÖR, Z.; MANNINGER, M.; JEREB, L. Application of sigmoid models for growth investigations of forest trees. Advances in Intelligent Systems and Computing, v. 282, p. 353–364, 2014.

PRODAN, M. et al. Mensura forestal. San José, Costa Rica.: GTZ/IICA-Agroamerica, 1997. 586 p.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (R. D. C. Team, Ed.): R Foundation for Statistical Computing.Vienna, AustriaR Foundation for Statistical Computing, 2019. Available at: http://www.r-project.org

RAPTIS, D. I. et al. Mixed-effects height–diameter models for black pine (*Pinus* nigra Arn.) forest management. Trees 2021 35:4, v. 35, n. 4, p. 1167–1183, 2021.

RICHARDS, F. J. A flexible growth function for empirical use. Journal of Experimental Botany, v. 10, n. 2, p. 290–301, 1959.

SCHRÖDER, J.; ÁLVAREZ, J. G. Comparing the performance of generalized diameter-height equations for Maritime pine in Northwestern Spain. Forstwissenschaftliches Centralblatt, v. 120, n. 1, p. 18–23, 2001. SCHUMACHER, F. X. A new growth curve and its application to timber yield studies. Journal of Forestry, v. 37, n. 819, p. 819–820, 1939.

SCHWARZ, G. Estimating the dimension of a model. The Annals of Statistics, v. 5, n. 2, p. 461–464, 1978.

SHARMA, M.; PARTON, J. Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. Forest Ecology and Management, v. 249, n. 3, p. 187–198, 2007.

SHARMA, R.; VACEK, Z.; VACEK, S. Nonlinear mixed effect height-diameter model for mixed species forests in the central part of the Czech Republic. Journal of Forest Science, v. 62, n. No. 10, p. 470–484, 2016.

SHARMA, R. P. et al. Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central Europe. Trees 2018 33:1, v. 33, n. 1, p. 103–119, 19, 2018.

VANCLAY, J. K. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests. 1. ed. Wallingford, U.K: CABI, 1994.

VARGAS-LARRETA, B. 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.

VONESH, E. F.; CHINCHILLI, V. M. Linear and Nonlinear Models for the Analysis of Repeated Measurements. NY, USA: Marcel Dekker Inc., 1997. 560 p.

WEIBULL, W. A Statistical Distribution Function of Wide Applicability. Journal of Applied Mechanics, v. 19, p. 293–297, 1951.

WEST, P. W.; RATKOWSKY, D. A.; DAVIS, A. W. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. Forest Ecology and Management, v. 7, n. 3, p. 207–224, 1984.

ZHANG, L. Cross-validation of non-linear growth functions for modelling tree height-diameter relationships. Annals of Botany, v. 79, n. 3, p. 251–257, 1997.