

## ARTÍCULOS

**Tree height prediction in Brazilian *Khaya ivorensis* stands**Predicción de la altura en plantaciones brasileñas de *Khaya ivorensis***Andressa Ribeiro \*\*, Antonio Carlos Ferraz Filho <sup>a</sup>, José Roberto Soares Scolforo <sup>b</sup>**

\* Corresponding author: <sup>a</sup> Universidade Federal do Piauí, Departamento de Engenharias - CPCE, Avenida Manoel Gracindo, Km 01, 64900-000, Bom Jesus, Piauí, Brasil, phone: +55 89 3562-1505, andressa.florestal@ufpi.edu.br

<sup>b</sup> Universidade Federal de Lavras, Departamento de Ciências Florestais, Campus Universitário, 37200-000, Lavras, Minas Gerais, Brasil.

## SUMMARY

Tree height measurement is one of the most difficult activities in forest inventory data gathering, although it is a fundamental variable to support forest management, since it is an input for modelling growth and yield. To overcome this obstacle and ensure that the heights of trees are estimated accurately, hypsometric relationships are used. Therefore, the objective of this study was to compare different fitting strategies (*i.e.* nonlinear least squares and mixed-effects) to predict tree height in African mahogany Brazilian plantations using well know local (using only tree height and diameter) and generalized (using height, diameter and plot level variables) models. Data were gathered on 149 permanent plots sampled in different Brazilian regions and ages, totaling 4,201 height-diameter pairs. Different models were evaluated and the best method to estimate the height-diameter relationship was based on statistical and graphical criteria. A local model using mixed-effects with correction of heteroscedasticity was efficient and superior to other models evaluated. However, when using an independent data base, the generalized model fitted by nonlinear least squares generates adequate results that are scaled to the plots' productivity, since the inclusion of dominant height into the model helps to predict height locally.

*Key words:* African mahogany, forest inventory, statistical modelling.

## RESUMEN

La medición de la altura del árbol es de difícil realización en inventarios forestales, aunque es una variable fundamental para apoyar el manejo forestal una vez que es dato de entrada para la modelación del crecimiento y producción. Para superar este obstáculo y garantizar un cálculo de alturas de los árboles con precisión se utiliza la relación hipsométrica. Por lo tanto, el objetivo de este estudio fue comparar diferentes estrategias de ajuste (mínimos cuadrados no lineales e efecto mixto) para predecir la altura de los árboles en plantaciones brasileñas de caoba africana (*Khaya ivorensis*) utilizando conocidos modelos locales (apenas diámetro e altura) y generalizados (diámetro, altura y variables de la parcela). Los datos fueron recogidos en 149 parcelas permanentes muestreadas en diferentes regiones brasileñas y edades, totalizando 4.201 pares de altura-diámetro. Diferentes modelos fueron evaluados y el mejor método para estimar la relación altura-diámetro se basó en los criterios estadísticos y gráficos. El modelo local usando efectos mixtos con la corrección de heterocedasticidad fue eficiente y superior a otros modelos evaluados. Sin embargo, cuando se utiliza una base de datos independiente, el modelo generalizado ajustado por mínimos cuadrados no lineales genera resultados adecuados que se ajustan a la productividad de las parcelas, ya que la inclusión de la altura dominante en el modelo ayuda a predecir la altura a nivel local.

*Palabras clave:* caoba africana, inventario forestal, modelización estadística.

## INTRODUCTION

Tree height measurement is one of the most difficult, time consuming and expensive activities in forest inventories data gathering (Ribeiro *et al.* 2010, Sharma and Breidenbach 2015); although it is a fundamental variable to support forest management since growth and yield models rely on accurate estimates of total tree height (Sharma

2016). In many situations, foresters save time and effort by measuring just a few trees inside the plot and predicting the other tree heights using a mathematical equation, highlighting the importance and widespread application of these models in forestry (Salas *et al.* 2010).

The height-diameter relationship (also known as hypsometric relationship) can be classified as local and regional/generalized (Trincado and Leal 2006, Paulo *et al.*

2011). Local models predict height variation using only one variable, commonly the diameter and are, thus, fitted by plot or stand. Regional models add, besides diameter, other stand variables that help explain height variation (e.g. dominant height, basal area, age) or even inclusion of random effects models and dummy variables for eco-region (Temesgen *et al.* 2014). Generalized models are fitted with larger databases than the ones used for local models and consequently are able to predict heights in diverse stand conditions.

Huang *et al.* (2009), Robinson and Hamann (2011), and Zang *et al.* (2016), among others, report that it is very common for data from natural resources to violate basic statistical assumptions for application of regression analysis via ordinary least squares method (OLS). This is because biological data may have temporal, spatial and hierarchical structure with dependence between observations, leading to biased estimate of the parameters' estimated variance. In mixed models, random effects are introduced in the model coefficients at different levels, such as region, site, plot and tree (Ou *et al.* 2015, Zang *et al.* 2016) to represent such dependences. Thus, studies using mixed-effects modeling have shown significant gains to predict tree height (Calama and Montero 2004, Shawn *et al.* 2009, Paulo *et al.* 2011).

African mahogany (*Khaya ivorensis* A. Chev.) cultivation is recent outside the species' countries of origin. In Brazil, the activity began in 1976 in the state of Pará and since then its cultivation has been an alternative to the native mahogany (*Swietenia macrophylla* G. King.) exploitation, since it is in threat of extinction (França *et al.* 2016) and restricted in homogeneous plantations due to attacks by the Meliaceae shoot borer *Hypsipyla grandella* (Zeller) (Lepidoptera: Pyralidae).

Due to the recent domestication of the species, few studies are conducted related to tree height prediction, es-

pecially in Brazil (Silva *et al.* 2016, Ribeiro *et al.* 2017). Therefore, the objective of this paper is to compare different fitting strategies to predict tree height in African mahogany Brazilian plantations using well know local and regional models fitted by: i) nonlinear least squares; ii) mixed-effects and iii) mixed-effects with correction of heteroscedasticity modelled by power-variance function. As a hypothesis, we expect that the modelling approach that most details the estimated values residual errors (*i.e.* mixed-effects with correction of heteroscedasticity modelled by a power-variance function) will yield the best results; also, we believe that regional models will provide a simple means for users to apply the models generated here to their own databases.

## METHODS

The data from the African mahogany plantations used in this research had similar forest management and genetic bases (Ribeiro *et al.* 2017); with age ranged from 2 to 14 years old and at least two re-measurements (table 1). The stands have most of soils classes belonging to latosol (Ribeiro 2017). In the state of Minas Gerais, four distinct types of climate predominate: Cwb, Cwa, Aw and BSw, according to Köppen climatic classification, with annual precipitations between 750 mm and 1,800 mm. In the state of Goiás the climate is tropical with dry season in the winter (Aw) and in the state of Pará the climate is tropical humid or superhumid (Af) or subhumid (Am) (Embrapa 2016).

Diameter (d) and height (h  $\geq$  1.3 m) data set were gathered using metric tape and Vertex III hypsometer, respectively, from 149 permanent plots located in different stands across Brazilian regions, totaling 4,201 height-diameter (h-d) pairs (figure 1).

**Table 1.** Characterization of the data set.

Caracterización de los datos.

Location/Coordinates	Data established	Total area (ha)	Plot area (m <sup>2</sup> )	Tree spacing (m)	Shape and number of plots	Number of remeasurements
Pirapora - MG/ 17°29'S 44°57'W	May/09	121.4	1,075	6x6	Rectangular (30)	5
Nova Ponte - MG/ 19°0.2'S 47°23'W	Nov/10	191.3	787	4x6	Rectangular (27)	3
São Roque de Minas - MG/ 20°6.6'S 46°27'W	Mar/10	56.5	800	5x7	Circular (27)	4
Piumhi - MG/ 20°25'S 46°1.3'W	Jan/09	9.2	900	5,5x6	Circular (14)	5
Nova Porteirinha - MG/ 15°46'S 43°18'W	Dec/10	241.6	736	4x6	Rectangular (10)	2
Nova Porteirinha - MG/ 15°39'S 43°17'W	Nov/08	96.0	728	4x6	Rectangular (8)	2
Rio Manso - MG/ 20°15'S 44°21'W	Apr/09	24.9	800	4x5	Circular (7)	2
Amorinópolis - GO/ 16°37'S 51°10'W	Feb/09	64.3	573	6x4	Circular (20)	2
Santo Antônio - PA/ 1°26'S 47°23'W	Jun/00	12.2	4,371	12x12	Rectangular (4)	2

Dominant height (hdom) was defined as being the mean height of the 30 thickest trees per hectare (Ribeiro *et al.* 2016), and not of the 100 thickest trees per hectare, as usual, due to the low density of the plantations in Brazil (spacing used at most plantations is 6 x 6 m, and 12 x 12 m for the oldest). Similar methodology was adopted by Paulo *et al.* (2011) for *Quercus suber* studies in Portugal and Danquah (2012) for African mahogany species planted in Ghana, where they selected 25 and 40 thickest trees per hectare, respectively, to determinate mean dominant height. The same procedure was done to determine dominant diameter (ddom), defined as being the mean diameter of the 30 thickest trees per hectare. We did not consider the effect of plot size and tree spatial distribution on dominant height and diameter estimation, though we expect any possible bias to arise from this to be neglectable. García and Batho (2005) reported mean bias values of 42 cm, given that the stands are homogeneous, and this effect is expected to be more important in more variable stands (García 1998).

A graphic of analysis of the data was performed for detection and exclusion of extreme observations, attributed to measurement errors, trees that were dead, damaged and presenting a broken top or trunk. A summary of the descriptive statistics of the data set used in this study is presented in table 2.

*Model fitting and selection.* In general terms, the regression analysis aims at representing the distribution of a response variable (Y) subject to values of a predictor variable of known values (X),  $f(Y|X_1, \dots, X_i)$ , as shown in [1].

$$Y_t = f(X_t; \theta) + \varepsilon_t \quad [1]$$

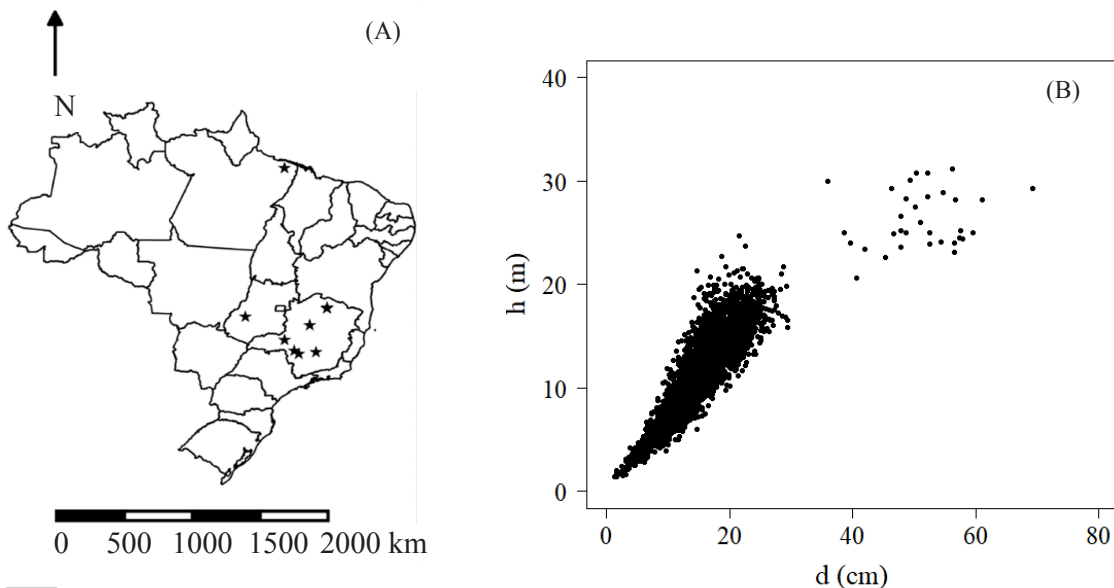
where:  $Y_t$  represents the response variable,  $X_t$  represents the predictor variable,  $\varepsilon_t$  represents the error term at the  $t$ th data point, and  $\theta$  is composed by a vector of the model's parameters.

Several models are used to represent the relationship between height and diameter in forest data and studies have emphasized the superiority of non-linear models (Huang *et al.* 2009, Mehtätalo *et al.* 2015). The mathematical expressions tested in this study are presented in table 3. A 1.3 constant was used to avoid the prediction of a height inferior to 1.3 meters when  $d$  is small.

**Table 2.** Summary of statistics for the African mahogany plantations hypsometric data set (n = 4,201), where SD = standard deviation.

Resumen estadístico del conjunto de datos usados para la hipsometría de plantaciones de caoba africano (n = 4.201).

Variable	Minimum	Mean	Maximum	SD
Diameter (cm)	1.3	15.1	69.3	5.7
Total height (m)	1.4	11.3	31.2	4.2
Dominant height (m)	3.8	12.3	28.2	3.9
Age (years)	1.9	4.2	14.0	1.6
Number of trees per hectare	55	320	563	70
Basal area (m <sup>2</sup> ha <sup>-1</sup> )	0.4	5.9	16.2	3.2



**Figure 1.** Stands location (represented by a star, A) and scatterplot between total tree height (h) and diameter at breast height (d) (n = 4,201, B).

Ubicación del rodal (representado por una estrella, A) y gráfico de altura total del árbol (h) contra el diámetro a la altura del pecho (d) (n = 4.201, B).

In table 3, models 1 to 8 are considered local, in which the height variation is estimated only based in the tree's diameter. Models 1 to 8 are traditionally used to describe the height-diameter relationship and were obtained from Mehtätalo *et al.* (2015). Models 9 to 15 are considered regional and insert additional stand variables to describe height variation. Models 9 to 15 were obtained from Scolforo (1990).

We compared local and regional models separately in this work. For each group of models (local and regional), a three-step fitting strategy was followed. The first

step consisted of model fitting without specifying any random effects, fitting a basic model by nonlinear least squares (NLS) techniques. All statistical inferences were made using the program R (R Core Team 2016) with the *nls* function performing a nonlinear regression analysis via Gauss Newton algorithm.

The second step (NLME) involved inclusion of random effects in the coefficients of the best models chosen in step 1, initially inserting random effects in all the coefficients of the models, as suggested by Pinheiro and Bates (2000), using the *nlme* package (Pinheiro *et al.* 2016) of R.

**Table 3.** Functions applied for height estimation\*.

Funciones utilizadas para la predicción de la altura\*.

Model	Mathematical function
1	$h = 1.3 + \frac{d^2}{(\beta_0 + \beta_1 d)^2}$
2	$h = 1.3 + \frac{\beta_0 d}{(1 + d)^{\beta_1}}$
3	$h = 1.3 + \frac{\beta_0 d}{\beta_1 + d}$
4	$h = 1.3 + \beta_0 \exp(\beta_1 d^{-1})$
5	$h = 1.3 + \frac{d^2}{\beta_0 d^2 + \beta_1 d + \beta_2}$
6	$h = 1.3 + \frac{\beta_0}{[1 + \exp(\beta_1 d / \beta_2)]}$
7	$h = 1.3 + \beta_0 [1 - \exp(-\beta_1 d)]^{\beta_2}$
8	$h = 1.3 + \beta_0 \exp[-\beta_1 \exp(-\beta_2 d)]$
9	$h = 1.3 + \exp \left[ \beta_0 + \beta_1 \left( \frac{1}{t} \right) + \beta_2 \ln(hdom) + \beta_3 \ln(G) \right]$
10	$h = 1.3 + \exp \left[ \beta_0 + \left( \frac{\beta_1}{d} \right) + \beta_2 \ln(hdom) \right]$
11	$h = 1.3 + \exp \left[ \beta_0 + \beta_1 \ln(hdom) + \beta_2 d^{-1} + \beta_3 \ln \left( \frac{N}{d} \right) + \beta_4 \cdot \ln \left( \frac{1}{td} \right) \right]$
12	$h = 1.3 + \exp \left[ \beta_0 + \beta_1 \ln(hdom) + \beta_2 t^{-1} + \beta_3 \ln(Nd) + \beta_4 \ln \left( \frac{1}{td} \right) \right]$
13	$h = 1.3 + (hdom - 1.3) \exp \left[ \left( \beta_0 - \beta_1 hdom + \beta_2 \frac{N}{1000} \right) \left( \frac{1}{d} - \frac{1}{d_{dom}} \right) \right]$
14	$h = 1.3 + \exp \left[ \beta_0 + \beta_1 t^{-1} + \beta_2 \ln(hdom) + \beta_3 \ln(G) + \beta_4 \ln \left( \frac{1}{td} \right) \right]$
15	$h = 1.3 + \exp \left[ \beta_0 + \beta_1 t^{-1} + \beta_2 \ln \left( \frac{dg}{d} \right) + \beta_3 \ln(dgt) \right]$

\*Where:  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$  = model parameters; h=total height (m); d=diameter at breast height (cm); hdom=dominant height (m); G=basal area ( $m^2 \cdot ha^{-1}$ ); N=number of trees per hectare; t=age (years); dg=quadratic mean diameter (cm) and ddom=dominant diameter (cm).

Coefficient estimation was based on the maximum likelihood and comparison of nested models tests were made based on the likelihood ratio (random part) and conditional F tests (fixed part). When mixed models are used, the goal is to predict values for Y from a continuous predictor variable X and add a categorical variable for each stipulated group. Following Calama and Monteiro (2004), Sharma and Parton (2007) and Pinheiro and Bates (2000), a general expression for a nonlinear mixed-effects model can be defined as [2].

$$Y_{ij} = f(\theta_i, X_{ij}) + \varepsilon_{ij} \quad [2]$$

where:  $Y_{ij}$  is the  $j$ th observation (tree) of the response variable taken from the  $i$ th sampling unit (plot),  $X_{ij}$  is the  $j$ th measurement from the predictor variable at the  $i$ th plot,  $\theta_i$  is a parameter vector, specific for each sampling unit,  $f$  is a nonlinear function of the predictor variables and the parameter vector, and  $\varepsilon_{ij}$  is the residual error term. In vector form, this mixed-effects model can be expressed as [3].

$$Y_i = f((A_i\beta + Z_i b_i), X_i) + \varepsilon_i \quad [3]$$

where:  $Y_i$  is the observation vector and  $X_i$  is the predictor vector for the  $i$ th plot,  $\varepsilon_i$  is a vector of the residuals,  $\beta$  is a vector of fixed population parameters,  $b_i$  is the vector of random-effects associated with the  $i$ th plot and  $A_i$  and  $Z_i$  are design matrices for the fixed- and random-effects specific to each plot, respectively.

We used as a random effect the combination of the local, plot and measurement occasion, totaling 380 groups. More detailed statistical notation and explanation of mixed modeling process can be found in Pinheiro and Bates (2000), Robinson and Hamann (2011) and Mehtätalo *et al.* (2015).

The third step (WNLME) was made when we verified violations of assumption of constant variance (homoscedasticity) in steps 1 and 2. If the plot of the standardized residuals versus the fitted values showed that the error variance was heterogeneous, a variance power function ( $\text{Var}(\varepsilon) = \sigma^2 (d)^{2\delta}$ ) was used to model the variance error structure within groups using covariates with an exponential parameter delta ( $\delta$ ) estimated by iterative processes. We chose initial  $\delta$  value = 0.5, which implies a linear relationship between the variance and tree diameter (Mehtätalo *et al.* 2015). A similar procedure was performed by Paulo *et al.* (2011) and Mehtätalo *et al.* (2015) to fit height-diameter models.

The models were chosen according to the goodness-of-fit, predictive ability, biological sense (e.g. positive height values, larger heights for larger diameter trees) and fulfillment of the error assumptions (homoscedasticity, lack of autocorrelation and normality of residuals). The goodness-of-fit of the functions was analyzed through the root mean squared error (RMSE, [4]) and value of the Akaike information criterion (AIC).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y - \hat{Y})^2}{n - 1}} \quad [4]$$

Where  $Y$  = observed response variable,  $\hat{Y}$  = estimated response variable and  $n$  = number of observations

## RESULTS

*Local models.* Statistical criteria and visual plot analysis of residuals versus fitted values for each fitted model (table 4) showed that the models 1, 4 and 7 had the best results considering the local models fitted by nonlinear regression using least squares method (NLS).

**Table 4.** Coefficients and goodness-of-fit of the local height diameter functions using least squares method (NLS)\*.

Coefficientes y bondades de ajuste de las funciones locales de altura y diámetro utilizando el método de los mínimos cuadrados (NLS)\*.

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	RMSE	RMSE%	AIC
1	2.504	0.145	-	1.78	15.8	16,767
2	0.914	0.109	-	2.28	20.2	17,712
3	105.022	139.178	-	1.93	17.1	17,445
4	32.732	-17.135	-	1.77	15.7	16,735
5	0.005	1.541	-2.146	2.06	18.2	17,990
6	23.021	15.147	6.889	1.81	16.0	16,919
7	25.946	0.065	1.966	1.76	15.6	16,675
8	23.550	3.561	0.096	1.78	15.8	16,754

\*All estimated parameters are significant at  $P$  level < 0.05.



The goodness-of-fit criteria for all equations were similar (table 4), with a slight superiority for model 7, followed by models 4, 8 and 1. Model 8 presented more coefficients than those shown by the others and was discarded in favor of a more parsimonious model. The residual plot for model 4 was biased, overestimating the predicted heights below 5 meters and all models showed trends of non-normality for higher values of prediction (figure 2).

The distribution of residuals for models 1 and 7 was similar, as was their goodness-of-fit, being model 1 chosen for the other two fitting strategies, since it has less parameters, and presented better fit for the higher height values (larger than 25 m). Proceeding to the second step of the fitting process, model 1 was fitted as a mixed-effects model with random effect inserted in all coefficients [5].

$$h_{ij} = 1.3 + \frac{d_{ij}^2}{[(\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) \cdot d_{ij}]^2} + \varepsilon_{ij} \quad [5]$$

where:  $u_{0i}$  and  $u_{1i}$  are the random-effects associated with parameters  $\beta_0$  and  $\beta_1$ , respectively.

The coefficients estimated for model 1 with the NLME and WNLME fitting strategies, the variance estimates for the random effects in the mixed model and the statistical criteria are presented in table 5.

The residual plots (figure 3) show a tendency of heterogeneity of the variance for the NLME method with

inclusion of random effects on the parameters, and this was corrected when using a power type variance function (WNLME) into the regression. Normality was not guaranteed for the extreme values of height prediction for both methodologies (figure 3).

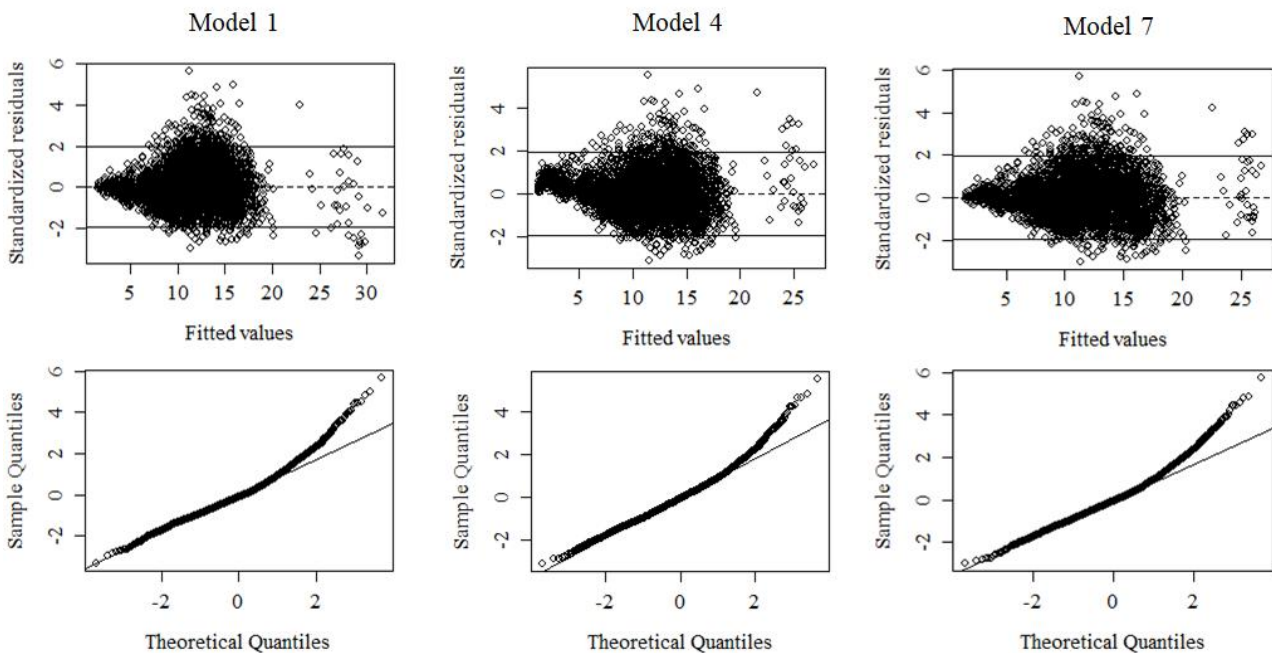
*Generalized models.* As for the local model fitting, generalized models were also first fit using the NLS method without hierarchy, resulting in the following best models: 10, 11 and 13 (table 6) and residuals plot shown in figure 4.

We chose the more parsimonious model (model 10) for fitting of the two other fitting strategies. The inclusion of the random effects in the model coefficient presented the best results, resulting in the following final model [6]:

$$h_{ij} = 1.3 + \exp \left[ (\beta_0 + u_{0i}) + \left( \frac{\beta_1}{d_{ij}} \right) + \beta_2 \cdot \ln(hdom_j) \right] + \varepsilon_{ij} \quad [6]$$

The coefficients estimated for model 10 with the NLME and WNLME fitting strategies, the variance parameters for the random effects in the mixed model and the statistical criteria are presented in table 7.

When fitting [6] with the inclusion of random effects, a minor improvement in the statistics used as selection criteria was observed, although the residual distribution presented similarity to the model without hierarchy (figure 4), with a slight bias for height prediction for trees under 5 meters (figure 5).



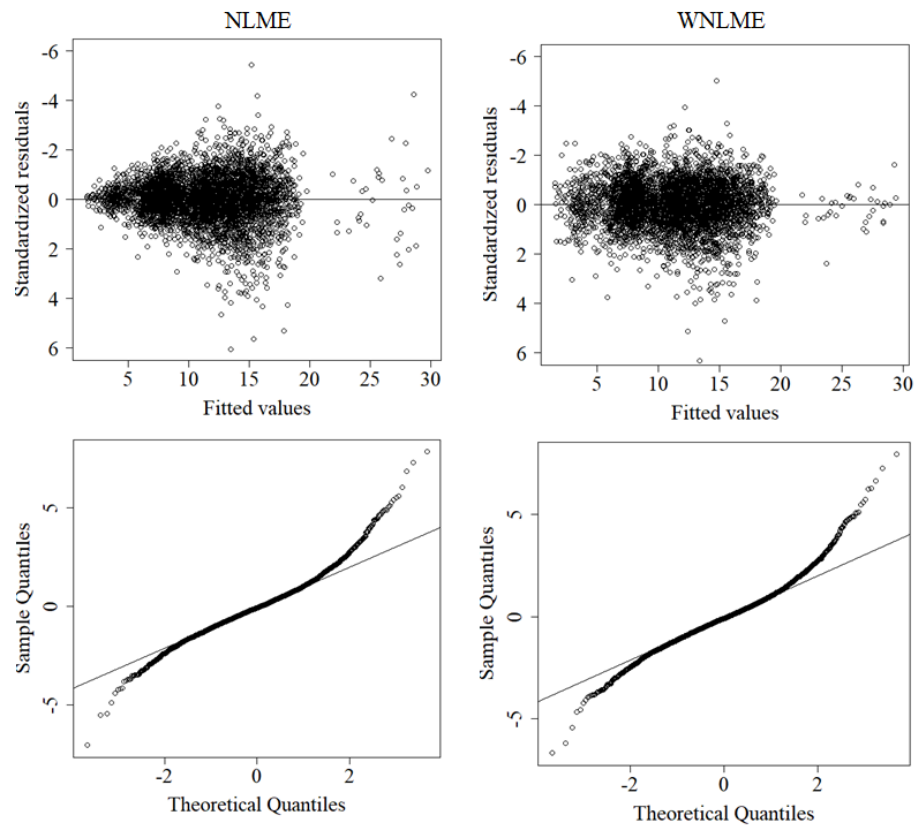
**Figure 2.** Residual versus fitted values and normal Q-Q plot for the best local fitted models.

Valores residuales versus valores ajustados y Q-Q normal gráfico para los mejores modelos locales ajustados.

**Table 5.** Estimated model coefficients and random effects, variance components for the random effects for Model 1, where SE = standard error and SD = standard deviation.

Estimativa de los coeficientes de lo modelo, efectos aleatorios, componentes de varianza para los efectos aleatorios para el modelo 1, donde SE = error estándar y SD = desviación estándar.

Strategy		Parameters	Value	SE	RMSE	AIC
2) NLME	Coefficients	$\hat{\beta}_0$	1.9759	0.0414	1.23	15,222
		$\hat{\beta}_1$	0.1801	0.0021		
	Random effects	SD ( $u_0$ )	0.6069			
		SD ( $u_1$ )	0.0224			
		Correlation: ( $u_0, u_1$ )	-0.8780			
		Residual variance	1.2991			
3) WNLME	Coefficients	$\hat{\beta}_0$	2.0561	0.0392	1.24	14,710
		$\hat{\beta}_1$	0.1756	0.0020		
	Random effects	SD ( $u_0$ )	0.5808			
		SD ( $u_1$ )	0.0221			
		Correlation ( $u_0, u_1$ )	-0.8630			
		Residual variance	0.1388			
		Variance power function	0.8219			



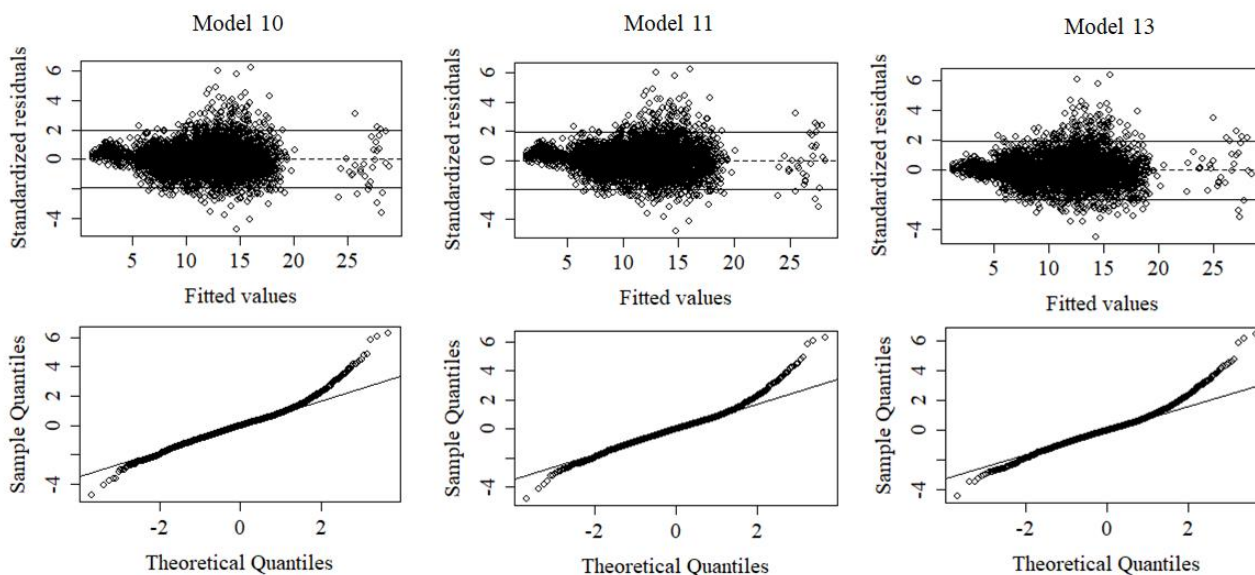
**Figure 3.** Residual versus fitted values and normal Q-Q plot for model 1 with different fitting strategies.

Valores residuales versus valores ajustados y Q-Q normal gráfico para el modelo 1 con diferentes estrategias de ajuste.

**Table 6.** Coefficients and goodness-of-fit of the generalized height diameter functions\*.  
 Coeficientes y bondades de ajuste de las funciones generalizadas de altura y diámetro\*.

NLS	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	RMSE	AIC
Model 9	0.0785	-0.6823	0.9246	0.0402 <sup>ns</sup>		1.92	17,399
Model 10	1.3209	-9.1651	0.6427			1.38	14,659
Model 11	1.4024	0.6252	-9.7249	0.0115	-0.0265 <sup>ns</sup>	1.38	14,652
Model 12	0.2828	0.6651	-2.2498	0.0231	0.5615	1.45	15,077
Model 13	-7.1662	0.0807 <sup>ns</sup>	-3.2597 <sup>ns</sup>			1.41	14,784
Model 14	0.5014	-2.3690	0.6695	-0.0149 <sup>ns</sup>	0.5792	1.46	15,085
Model 15	1.5972	-1.6892	-0.6270	0.2775		1.74	16,576

\*ns = estimated parameters not significant at  $P$  level  $> 0.05$ .



**Figure 4.** Residual versus fitted values and normal Q-Q plot for the best generalized models fitted.

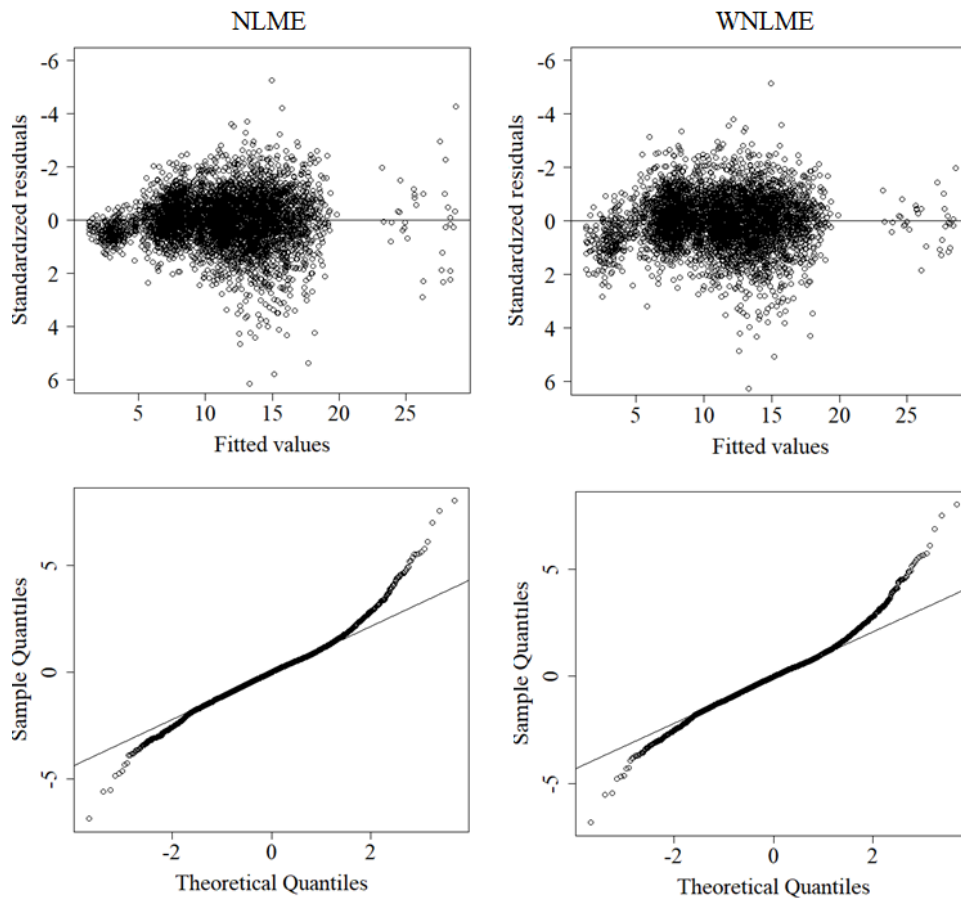
Valores residuales versus valores ajustados y Q-Q normal gráfico para los mejores modelos generalizados ajustados.



**Table 7.** Estimated coefficients and random effects, variance components for random effects for model 10, where SE = standard error and SD = standard deviation.

Estimativa de los coeficientes de lo modelo, efectos aleatorios, componentes de varianza para los efectos aleatorios para el modelo 10, donde SE = error estándar y SD = desviación estándar.

Strategy	Parameters	Value	SE	RMSE	AIC	
2) NLME	Coefficients	$\hat{\beta}_0$	1.3760	0.0469	1.28	14,493
		$\hat{\beta}_1$	-9.3996	0.1790		
		$\hat{\beta}_2$	0.6268	0.0149		
	Random effects	SD ( $u_0$ )	0.0427			
		Residual variance	1.3035			
3) WNLME	Coefficients	$\hat{\beta}_0$	1.5050	0.0441	1.27	14,187
		$\hat{\beta}_1$	-9.5485	0.1504		
		$\hat{\beta}_2$	0.5807	0.0145		
	Random effects	SD ( $u_0$ )	0.0508			
		Residual variance	0.2789			
		Variance power function	0.5688			



**Figure 5.** Graphical relationship between the standardized residuals and fitted values for model 10 with different fitting strategies.

Relación gráfica entre los valores de los residuos estandarizados y los valores estimados para el modelo 10 con diferentes estrategias de ajuste.

## DISCUSSION

The main objective of the present study was to develop equations that adequately predicted height for African mahogany stands in Brazil, respecting statistical assumptions and parsimony. While some studies have reported growth parameters and wood quality for *Khaya ivorensis* plantations considering limited stand variations (e.g. Silva *et al.* 2016, Ribeiro *et al.* 2016), to our knowledge, this study is the first to provide height-diameter models for a large scale database in Brazil and beyond. Care must be taken when applying the models outside the sampled database range (for other parts of the world or for ages over 14 years), especially considering the peculiarities of Brazilian African mahogany silviculture (intensive management practices and wide spacing).

It is expected that a model including stand variables (i.e. generalized models) provide a better predictive equation for height, as noted by Trincado and Leal (2006). It was clear in this work that when the models were fitted by NLS method, a predictive improvement of a half meter error comparing local model 1 with generalized model 10 occurred, besides the lowest AIC value for the last equation.

Mixed-effects modeling is one alternative to deal with correlated observations, in which the variability between the sampling units can be explained by including random effects, which are estimated at the same time as the model coefficients (Calama and Montero 2004). Sharma and Parton, (2007) developing a h-d equation for species of boreal forest in Canada, found that the inclusion of random effects on the selected model coefficients resulted in a lower value of AIC and an improvement in the models' predictive ability. Temesgen *et al.* (2014), fitting the ratio of height and diameter for various species in China, used the mixed-effects method to correct the hierarchical data structure and generate a robust predictive equation. Our results confirmed this trend, where the mixed-effects models provided better results compared to the NLS techniques.

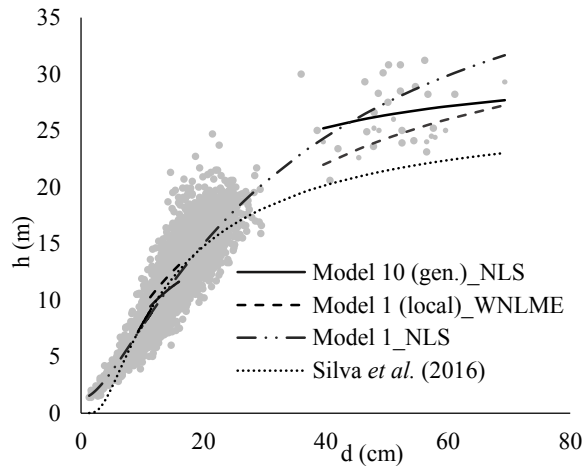
For all selected models non-normality for extreme values occurred. Zang *et al.* (2016) affirmed that since the height-diameter relationship is influenced by numerous factors, it may be difficult to model using normal parametric models and limitation of least-squares methods (e.g. normally distributed errors) may present problems, especially in the case of generalized h-d equations. Crecente-Campo *et al.* (2010) suggest the use of weighting factors to balance error variance, to account for non-normality and to take into account unequal selection probabilities. Although the impact of the weighting procedure was minimal in their work, the parameter estimates and approximate standard errors showed the same magnitude, the goodness-of-fit statistics was also similar, with slightly better values for the model fitted using unequal selection probabilities.

For the selected generalized model (model 10), the inclusion of a random effect did not result in explicit improvement of residual distribution (figure 5), with slight improvement on statistics values (table 7) compared with the NLS method. That was expected since the inclusion of a stand variable into the model works as a plot level control, improving the predictions in local scale. The small effect of the random component for the generalized model was confirmed by its low value of standard deviation, 0.04 m, compared with a residual standard deviation of 1.14 m for the NLME fit (table 7). The random component's standard deviation values were much more expressive for the local models, reaching 0.61 m, compared with the residual standard deviation of 1.14 m for the NLME fit (table 5). However, when heteroscedasticity was corrected, residuals were less biased and the values of the statistics were higher than those for the other fitting strategies. The relationship defined between the standardized residuals and the tree height estimates did not suggest the presence of heteroscedasticity associated with the error term for WNLME approach in the local and generalized model selected, although non-normality still existed. Calegario *et al.* (2005), estimating the height growth of clonal *Eucalyptus* trees, obtained significant gains when modeling the heterogeneity of variance, where the distribution of residues was significantly improved. We also arrived at the same conclusion when we applied the variance power function on the selected models.

In the present study, the gain in the use of a generalized model (using dominant height) compared with local models including random effects on the parameters was not very significant. It is known that the dominant height is a variable that reflects local productivity, being correlated with the total height of the trees; hence, the inclusion of the same in hypsometric designs results in improvement of height predictions.

We plotted the different height-diameter selected models in varying silvicultural scenarios (one older plot with large spacing and one younger plot with dense spacing) to illustrate the models' predictive ability, and also to highlight the errors of selecting equations from literature without any calibration (figure 6). In figure 6 we see that the most general model from our study (Model 1\_NLS) is able to adequately describe the mean behavior of the height-diameter relationship, nonetheless it is not able to distinguish between sites showing different productivity. Considering the strategies that take into account the plots' productivity (Model 10 (gen.) NLS and Model 1 (local) WNLME), we see that they are able to better describe particular differences in the height-diameter relationship in different plots (represented as black dots in figure 6), with superiority for the mixed-effects approach.

Thus, the authors suggest that when a large-scale database is available to fit height-diameter models for *Khaya ivorensis*, model 1 using random location parameters and correction of the heteroscedasticity of the residuals should



**Figure 6.** Different height predictions for African mahogany dataset, where the gray dots show the original data and the dark dots show the observed heights for two sample plots.

Diferentes predicciones de altura para el conjunto de datos de caoba africano, donde los puntos grises muestran los datos originales y los puntos oscuros muestran las alturas observadas para dos parcelas.

be the preferred method to estimate the height of unsampled trees. However, model users that do not have these data can use the generalized model (model 10), since inclusion of the dominant height into the model helps to predict height locally. For Salas *et al.* (2016), in general, models are good for the purpose they were built for and it is very difficult to find a model that works well for all purposes.

Finally, to illustrate the need for calibration of the height-diameter relationship and to adhere to the range of the sampled database, we also plotted a model from Silva *et al.* (2016) in figure 6. These authors presented a local equation fitted with NLS method to predict height in 4-year-old African mahogany stands geographically close to some of the stands used in this study. Here we see that while Silva's model predicted the height variation reasonably well for trees with diameters with 10 to 20 cm, it provided large underestimation outside this range, especially for larger trees ( $d > 40$  cm).

## CONCLUSIONS

We concluded that the modelling approach that most details the estimated height values residual errors (mixed-effects considering each plot measurement occasion as the random effect and with correction of heteroscedasticity modelled by power-variance function) yielded the best results. However, given the fact that users will not always have a large database at their disposal, the generalized model 10 fitted using the plot dominant height can be applied to successfully estimate tree height of other *Khaya ivorensis* stands.

## REFERENCES

- Calama R, G Montero. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Canadian Journal of Forestry Research* 34: 150-163. DOI: 10.1139/x03-199.
- Calegario N, RE Daniels, R Maestri, R Neiva. 2005. Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. *Forest Ecology and Management* 204: 11-20. DOI: 10.1016/j.foreco.2004.07.051.
- Crecente-Campo F, M Tomé, P Soares, U Diéguez-Aranda. 2010. A generalized nonlinear mixed-effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. *Forest Ecology and Management* 259: 943-952. DOI: 10.1016/j.foreco.2009.11.036.
- Danquah JA. 2012. Restoration of degraded dry semideciduous forest ecosystems in Ghana: Effects of African mahogany species on soil chemistry, tree diversity and the application of leaf morphometrics for provisional seed zonation. Dissertation in Forest Sciences. Joensuu, Finland. Faculty of Science and Forestry, University of Eastern Finland. 37 p.
- Embrapa. Clima. Consulted 15 mar. 2016. Available in <https://www.cnpf.embrapa.br/pesquisa/efb/clima.htm>
- França TSFA, FJN França, RA Arango, BM Woodward, MDC Arantes. 2016. Natural resistance of plantation grown African mahogany (*Khaya ivorensis* and *Khaya senegalensis*) from Brazil to wood-rot fungi and subterranean termites. *International Biodeterioration & Biodegradation* 107: 88-91. DOI: 10.1016/j.ibiod.2015.11.009.
- García O. 1998. Estimating top height with variable plot sizes. *Canadian Journal of Forest Research* 28(10): 1509-1517. DOI: 10.1139/x98-128.
- García O, A Batho. 2005. Top height estimation in lodgepole pine sample plots. *Western Journal of Applied Forestry* 20(1): 64-68.
- Huang S, SX Meng, Y Yang. 2009. Using nonlinear mixed model technique to determine the optimal tree height prediction model for Black Spruce. *Modern Applied Science* 3(4): 3-18. DOI: 10.5539/mas.v3n4p3.
- Mehtätalo L, S de-Miguel, TG Gregoire. 2015. Modeling height-diameter curves for prediction. *Canadian Journal of Forest Research* 45:826-837. DOI: 10.1139/cjfr-2015-0054.
- Ou G, J Wang, H Xu, K Chen, H Zheng, B Zhang, X Sun, T Xu, Y Xiao. 2015. Incorporating topographic factors in nonlinear mixed-effects models for aboveground biomass of natural Simao pine in Yunnan, China. *Journal of Forestry Research* 27(1):119-131. DOI: 10.1007/s11676-015-0143-8.
- Paulo JA, J Tomé, M Tomé. 2011. Nonlinear fixed and random generalized height-diameter models for Portuguese cork oak stands. *Annals of Forest Science* 68:295-309. DOI: 10.1007/s13595-011-0041-y.
- Pinheiro J, D Bates. 2000. Mixed-effects models in S and Splus. New York, USA. Springer-Verlag. 528 p.
- Pinheiro J, D Bates, S DebRoy, D Sarkar and R Core Team. 2016. nlme: Linear and nonlinear mixed effects models. R Package Version 3.1-125. Consulted 15 mar. 2016. Available at <https://cran.r-project.org/web/packages/nlme/nlme.pdf>.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

- Vienna, Austria. Consulted in 10 feb. 2016. Available at <http://www.R-project.org/>
- Ribeiro A, AC Ferraz Filho, JM Mello, MZ Ferreira, PMM Lisboa, JRS Scolforo. 2010. Estratégias e metodologias de ajuste de modelos hipsométricos em plantios de *Eucalyptus* sp. *Cerne* 16(1): 22-31. DOI: 10.1590/S0104-77602010000100003.
- Ribeiro A, AC Ferraz Filho, M Tomé, JRS Scolforo. 2016. Site quality curves for African mahogany plantations in Brazil. *Cerne* 22(4): 439-448. DOI: 10.1590/01047760201622042185.
- Ribeiro A, AC Ferraz Filho, JRS Scolforo. 2017. O cultivo do mogno africano e o crescimento da atividade no Brasil. *Floresta e Ambiente* 24: e00076814. DOI: 10.1590/2179-8087.076814.
- Ribeiro A. 2017. African mahogany plantations: modelling growth and yield in Brazil. Tese em Ciências Florestais. Lavras, Brasil. Universidade Federal de Lavras. 165 p.
- Robinson AP, JD Hamann. 2011. Forest analytics with R: an introduction. New York, USA. Springer-Verlag. 354 p.
- Salas C, L Ene, TG Gregoire, E Næsset, T Gobakken. 2010. Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment* 114(6): 1277-1285. DOI: 10.1016/j.rse.2010.01.020
- Salas C, TG Gregoire, DJ Craven, H Gilibert. 2016. Modelación del crecimiento de bosques: estado del arte. *Bosque* 37(1): 3-16. DOI: 10.4067/S0717-92002016000100001
- Scolforo JRS. 1990. Sistema integrado para predição e análise presente e futura do crescimento e produção, com otimização de remuneração de capitais, para *Pinus caribaea* var. *handurensis*. Tese em Engenharia Florestal. Curitiba, Brasil. Universidade Federal do Paraná. 290 p.
- Sharma M, J Parton. 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management* 30: 187-198. DOI: 10.1016/j.foreco.2007.05.006
- Sharma M, J Breidenbach. 2015. Modeling height-diameter relationships for Norway spruce, Scots pine, and downy birch using Norwegian national forest inventory data. *Forest Science and Technology* 11(1): 44-53. DOI: 10.1080/21580103.2014.957354.
- Sharma M. 2016. Comparing height-diameter relationships of boreal tree species grown in plantations and natural stands. *Forest Science* 62(1): 70-78. DOI: 10.5849/forsci.14-232.
- Shawn X, SX Meng, S Huang, Y Yang, G Trincado, CL Vanderschaaf. 2009. Evaluation of population-averaged and subject-specific approaches for modeling the dominant or codominant height of lodgepole pine trees. *Canadian Journal of Forest Research* 39(6): 1148-1158. DOI: 10.1139/X09-039.
- Silva LF, GL Ferreira, ACA Santos, HG Leite, ML Silva. 2016. Equações hipsométricas, volumétricas e de crescimento para *Khaya ivorensis* plantada em Pirapora. *Floresta e Ambiente* 23(3): 362-368. DOI: 10.1590/2179-8087.130715.
- Temesgen H, CH Zhang, XH Zhao. 2014. Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China. *Forest Ecology and Management* 316: 78-89. DOI: 10.1016/j.foreco.2013.07.035.
- Trincado G, C Leal. 2006. Ecuaciones locales y generalizadas de altura-diámetro para pino radiata (*Pinus radiata*). *Bosque* 27(1): 23-34. DOI: 10.4067/S0717-92002006000100003.
- Zang H, X Lei, W Zeng. 2016. Height-diameter equations for larch plantations in northern and northeastern China: a comparison of the mixed-effects, quantile regression and generalized additive models. *Forestry* 89: 434-445. DOI: 10.1093/forestry/cpw022.

Recibido: 17.01.17

Aceptado: 31.10.17