For. Sci. 60(3):560–568
http://dx.doi.org/10.5849/forsci.12-099

biometrics

A Comparison of Model Forms for the Development of Height-Diameter Relationships in Even-Aged Stands

Esteban Gómez-García, Ulises Diéguez-Aranda, Fernando Castedo-Dorado, and Felipe Crecente-Campo

Several model forms were analyzed to develop a height-diameter (*h-d*) relationship for even-aged, birch-dominated stands (8,454 *h-d* pairs of 198 plot-inventory combinations) in northwestern Spain. A basic model (which includes only *d* as predictor) and a generalized model (which also considers stand variables as regressors) were selected from among available models. Fixed and mixed forms of these models were evaluated. An approximate Bayesian estimator was used to obtain a calibrated response (prediction with fixed parameters and random effects) of the mixed models from prior height measurements of trees selected by different strategies. From a practical point of view, the calibrated basic mixed model is recommended if a randomly selected sample of 4–11 trees is available or if only the 3 trees that are the diameter quartiles are measured. If the random sample includes more than 11 trees, the basic fixed model should be locally fitted, and the generalized fixed model is recommended when only dominant height is known. Finally, the predictions of the latter model may be outperformed by the calibrated generalized mixed model, which also requires inclusion of the height of the smaller diameter trees.

Keywords: birch, *Betula pubescens* Ehrh., mixed models, calibration, Galicia

The most important individual tree variables in forest management are dbh (*d*) and total tree height (*h*). In field inventories, the cost of measuring *h* is greater than that of measuring *d*, which favors the use of equations to estimate *h* from measured *d*. Tree height is important in the study of growth and yield because it is used to characterize the stand structure, to estimate individual tree and stand volumes, and to determine dominant height and site index (Peng 2001).

In even-aged stands, the differences in *h-d* relationships mainly depend on age, density, and site (Curtis 1967, Prodan et al. 1997). Therefore, a single curve cannot be used to estimate all possible *h-d* relationships that may be found within a forest. The method most commonly used to minimize this level of variance is to fit a "basic model" (which includes dbh as the sole independent variable) for each plot and measurement occasion. This modeling approach represents the best equation that can be obtained for a specific stand and point in time. The main problem with this approach is the large sampling effort required. Generalized and mixed *h-d* models are alternative approaches that reduce this effort.

"Generalized models" use *d* and stand-specific variables as regressors, which account for the differences in the *h-d* relationship across stands and within stands over time. Several generalized and region-wide equations have been developed for many tree species (e.g., Curtis 1967, Larsen and Hann 1987, López Sánchez et al. 2003, Temesgen and Gadow 2004).

Mixed models allow for both *mean* and *subject-specific* responses. The first considers only fixed parameters, common to the population, whereas the second considers both fixed parameters and random effects, specific to each subject (in this study each plot-inventory combination). The inclusion of random effects enables the variability in the *h-d* relationship between different locations and time to be modeled, after defining a common fixed functional structure (Lindstrom and Bates 1990). If a prediction for a new stand at a specific point in time is required and prior information is available (a sample of trees measured for *h* and *d*), the random effects of the *h-d* curve can be predicted, which provides for a *calibrated* response. Several studies have used mixed models to develop *h-d* relationships (e.g., Lappi 1997, Calama and Montero 2004, Sharma and Parton 2007, Trincado et al. 2007).

Birch (*Betula pubescens* Ehrh., also referred to as *Betula alba* L. and *Betula pubescens* subsp. *celtiberica* Rothm. & Vasc.: Castroviejo et al. 1990) is an important fast-growing pioneer tree species in Galicia (northwestern Spain), where it covers 32,000 ha as the main tree species (Xunta de Galicia 2001). Management of broadleaf

Manuscript received August 14, 2012; accepted June 25, 2013; published online August 29, 2013.

Affiliations: *Esteban Gómez-García (esteban.gomez@usc.es), University of Santiago de Compostela, Departamento de Ingeniería Agroforestal, Escuela Politécnica Superior, Benigno Ledo, Lugo, Spain. Ulises Diéguez-Aranda (ulises.dieguez@usc.es), University of Santiago de Compostela. Fernando Castedo-Dorado (fcasd@unileon.es), University of Leon. Felipe Crecente-Campo (felipe.crecente@usc.es), University of Santiago de Compostela.*

Acknowledgments: *This work was supported by the Spanish Ministry of Education and Science (AGL2007-66739-C02-01) and the European Regional Development Fund.*

Table 1. Summary statistics of the fitting data set (8,454 *h-d* observations in 198 plot-inventory combinations).

Variable	Mean	Minimum	Maximum	SD
d (cm)	15.3	4.7	42.4	5.6
h (m)	14.3	4.3	27.0	3.7
t (yr)	35.2	12	94	11.9
N (trees ha^{-1})	1,707	350	6,000	1,048
d_g (cm)	15.3	7.4	25.1	3.6
G (m^2 ha^{-1})	27.6	3.3	71.8	10.9
$\bar{h}$ (m)	14.3	6.5	20.0	2.9
d_0 (cm)	22.6	10.7	36.6	4.9
H (m)	16.5	7.2	24.5	3.6

d, diameter at breast height (1.3 m above ground level) outside bark; h, total tree height; t, stand age; N, number of live trees per hectare; d_g, quadratic mean diameter; G, stand basal area; $\bar{h}$, mean height; d_0, dominant diameter, defined as the mean diameter of the 100 largest-diameter trees per hectare; H, dominant height, defined as the mean height of the 100 largest-diameter trees per hectare.

species, including birch, is important for preserving biodiversity and also from an economic point of view. Despite this, there is a lack of management tools, such as *h-d* relationship, for this species in the region.

The overall objective of the present study was to compare several *h-d* model forms for even-aged, birch-dominated stands in Galicia. The two specific objectives were as follows: to evaluate alternative fixed and mixed, basic, and generalized *h-d* models and to compare these models in terms of predictive capability and sampling effort, thus providing recommendations for practical use.

Data

The data were obtained from a network of research plots installed in even-aged, birch-dominated stands (85% or more of the standing basal area consisting of birch) throughout the area of distribution of this species in Galicia. In the winters of 1997–1998, 1998–1999, and 2001–2002, 137 plots were subjectively selected to represent the existing range of ages, stand densities, and sites. The plot size ranged from 200 to 1000 m^2, depending on stand density. Although the smallest plot size may appear very small, only 2 plots (selected to include extreme site and stand conditions) were of this size. The other plots were larger than 500 m^2.

All trees in each sample plot were labeled. Two measurements of dbh (1.3 m aboveground level) were made at right angles to each other (with tree calipers, to the nearest 0.1 cm) and the arithmetic mean value was calculated. Total height was measured with a hypsometer, to the nearest 0.25 m, in a randomized sample of 30 trees and in an additional sample including the dominant trees (the proportion of the 100 largest-diameter trees per hectare, depending on plot size). Descriptive variables for each tree were also recorded, e.g., if they were alive or dead.

All of the initially established plots that were not lost because of forest fires or clear cutting were remeasured in the winter of 2008–2009 (50 plots). In addition, 11 new plots were established in the winter of 2009–2010 to complete the database for this and future studies. In both of these inventories, the total height of all trees was measured with a digital hypsometer, to the nearest 0.1 m, together with dbh and descriptive variables.

To develop the generalized model, different stand variables were calculated for each of the 198 plot-inventory combinations (first inventory of 148 plots plus second inventory of 50 plots). Summary statistics of the individual tree and stand variables used to fit the models are shown in Table 1.

Methodology

Models

In an initial step, five basic models (four already analyzed by Calama and Montero [2004] for stone pine in Spain and a logarithmic model) were fitted by ordinary nonlinear least squares (ONLS) for each plot-inventory combination. The model developed by Burkhart and Strub (1974) performed slightly better than the others and was selected as the basic model

$$h_{ij} = 1.3 + (a_1 + u_{1i})\exp\left(\frac{a_2 + u_{2i}}{d_{ij}}\right) + e_{ij} \qquad (1)$$

where h_{ij} is the total height of the jth tree in the ith plot-inventory combination, d_{ij} is its dbh, a_k are fixed parameters ($k = 1, 2$), u_{ki} are the corresponding random effects, and e_{ij} are tree-level errors independent of the u_{ki}.

Two different sets of random effects assumptions were investigated, leading to the following model forms

Model I: u_{ki} are fixed with $u_{ki} = 0$; $\mathrm{Var}(e_{ij}) = \sigma^2$.

Model II: u_{ki} are random with $\mathrm{E}[u_{ki}] = 0$, $\mathrm{Var}(u_{ki}) = \sigma^2_{u_k}$,

and $\mathrm{Cov}(u_{1i}, u_{2i}) = \sigma_{u_1u_2}$; $\mathrm{Var}(e_{ij}) = \sigma^2$.

In a second step, several generalized models selected from previous studies (Krumland and Wensel 1988, Tomé 1988, López Sánchez et al. 2003, Sharma and Zhang 2004, Castedo Dorado et al. 2006, Sharma and Parton 2007, Crecente-Campo et al. 2010) were fitted by ONLS to the whole data set. A new generalized model was also developed from Equation 1, by relating its fixed parameters to stand variables (graphical analysis was used to investigate the type of relationship: linear, power function, or exponential) and imposing the restriction to pass through the point (dominant diameter [d_{0i}] and dominant height [H_i]). The generalized form of Equation 1 performed slightly better than the other candidate models and was selected as the generalized model

$$h_{ij} = 1.3 + (H_i - 1.3)\exp\left(((a_1 + u_{1i}) + (a_2 + u_{2i})H_i + (a_3 + u_{3i})d_{gi})\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right) + e_{ij} \qquad (2)$$

where d_{gi} is the quadratic mean diameter of the ith plot-inventory combination, a_k are fixed parameters ($k = 1, 2, 3$), and u_{ki} are the corresponding random effects; the remaining variables were defined previously.

Again, two different sets of random effects assumptions were investigated, leading to the following model forms:

Model III: u_{ki} are fixed with $u_{ki} = 0$; $\mathrm{Var}(e_{ij}) = \sigma^2$.

Model IV: u_{ki} are random with $\mathrm{E}[u_{ki}] = 0$, $\mathrm{Var}(u_{ki}) = \sigma^2_{u_k}$,

$\mathrm{Cov}(u_{1i}, u_{2i}) = \sigma_{u_1u_2}$, $\mathrm{Cov}(u_{1i}, u_{3i}) = \sigma_{u_1u_3}$,

and $\mathrm{Cov}(u_{2i}, u_{3i}) = \sigma_{u_2u_3}$; $\mathrm{Var}(e_{ij}) = \sigma^2$.

The mixed models II and IV include both fixed parameters (common to the population) and Gaussian-distributed random (subject-specific) effects (Lindstrom and Bates 1990, Pinheiro and Bates 1995). In accordance with Pinheiro and Bates (1995) and Fang and

Bailey (2001), an attempt was made to include random effects in all fixed parameters.

The inclusion of random effects was only considered at the plot-measurement level. Although some studies (e.g., Lappi 1997, Mehtätalo 2004) have dealt with the temporal autocorrelation resulting from repeated measurements in the development of *h-d* curves, this measurement-occasion correlation was not considered in the present study because the number of measurement occasions per plot (1.34 as average) was too small.

In models II and IV, the fixed parameters can be used to predict a mean response (i.e., setting the random effects to 0). If the vector of random effects ($\mathbf{u}_i$) is predicted for a new plot-inventory combination *i,* through a calibration process, the total height for each diameter can be estimated by using a specific calibrated model (Vonesh and Chinchilli 1997). Prediction of the vector of random effects requires one or more measurements of total tree height and can be done by use of an approximate Bayesian estimator (Vonesh and Chinchilli 1997)

$$\hat{\mathbf{u}}_i \cong \hat{\mathbf{D}}\mathbf{Z}_i^T(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^T + \hat{\mathbf{R}}_i)^{-1}\hat{\mathbf{e}} \tag{3}$$

where $\hat{\mathbf{D}}$ is the estimated variance-covariance matrix for the random effects $\mathbf{u}_i$, $\hat{\mathbf{R}}_i$ is the estimated variance-covariance matrix for the error term, $\mathbf{Z}_i$ is the partial derivatives matrix with respect to the random effects, and $\hat{\mathbf{e}}$ is the error matrix estimated using only the fixed parameters.

For the mixed models II and IV, the calibrated response was evaluated using different sampling designs (e.g., Castedo Dorado et al. 2006, Crecente-Campo et al. 2010):

1. Total height of randomly selected trees (1–15 in the basic mixed model and 1–10 in the generalized mixed model). The mean and extreme values of the root mean squared error (RMSE) after 100 simulations were obtained.
2. Total height of the trees used to calculate the dominant height (H_i) (only for the basic mixed model).
3. Total height of the 1–3 smallest diameter trees.
4. Total height of the 3 trees that are the diameter quartiles (*Q*).
5. Total height of the 4 trees that are the diameter quintiles (*K*).

An example of the calibration process for both basic and generalized mixed models is shown in the Appendix.

Model I was also refitted by plot-inventory data with measurements from 5–15 randomly selected trees. The mean and extreme values of the RMSE after 100 simulations were obtained. Smaller samples were not considered because previous analyses showed that fitting the basic fixed model with less than 5 trees (regardless of the sampling design used) produces large errors.

Table 2. RMSEs of the basic fixed model I fitted by ONLS for each plot-inventory combination with all observations and observations for 5–15 randomly selected trees.

Data used in local fit	RMSE (m)
All observations	1.532
5 random trees	3.674 (1.535–30.69)
6 random trees	2.071 (1.534–8.483)
7 random trees	1.855 (1.534–4.929)
8 random trees	1.785 (1.534–4.204)
9 random trees	1.736 (1.533–3.406)
10 random trees	1.700 (1.533–2.780)
11 random trees	1.676 (1.533–2.530)
12 random trees	1.657 (1.533–2.343)
13 random trees	1.644 (1.533–2.285)
14 random trees	1.632 (1.533–2.165)
15 random trees	1.621 (1.533–2.080)

Mean and minimum and maximum (in parentheses) RMSE values after 100 fittings are shown.

Model Fit and Selection

The basic and generalized fixed models (I and III, respectively) were fitted by ONLS, with the NLIN procedure of SAS/STAT (SAS Institute, Inc. 2009), whereas the basic and generalized mixed models (II and IV, respectively) were fitted using the maximum likelihood estimation by the adaptive Gauss-Hermite quadrature method implemented in the NLMIXED procedure of the same statistical analysis software. Evaluation of the different models was based on statistical and graphical analysis of the residuals.

Two statistical criteria obtained from the residuals were examined: the model efficiency (EF), which is similar to the coefficient of determination for linear regression and indicates the proportion of the total variance that is explained by the model, and the RMSE. The RMSE is useful because it is expressed in the same units as the dependent variable, which therefore provides an idea of the mean error involved in using the model. Moreover, the RMSE penalizes models with more parameters, according to the general scientific principle of simplicity (Peña 2002). The RMSE was also used to compare the different fitting methodologies and calibration options evaluated, because the EF value in mixed models analysis does not connote "percent variation explained," as is usually understood and often leads to negative or meaningless values, even for moderately well-fitted models (Huang et al. 2009). The same can be applied to locally fitted basic fixed models. The expressions of these statistics are summarized as follows

$$EF = 1 - \frac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n_i}(y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n_i}(y_{ij} - \bar{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n_i}(y_{ij} - \hat{y}_{ij})^2}{n - p}} \tag{5}$$

where y_{ij} and $\hat{y}_{ij}$ are the *j*th observed and predicted values in the plot-inventory combination *i,* respectively, $\bar{y}$ is the grand mean of the observed values, n_i is the number of trees in the *i*th plot-inventory combination, *m* is the number of plot-inventory combinations, *n* is the number of observations used to fit the model, and *p* is the number of model parameters.

These goodness-of-fit statistics were computed over the complete data set, regardless of which trees were used for fitting or calibration. This procedure may lead to goodness-of-fit statistics that will be biased upward (EF) or downward (RMSE) because the same trees (and plots in the case of models III and IV) are used to estimate the fixed parameters (and the variance components for the random effects of models II and IV) and to evaluate the different model fits or calibrations. The magnitude of this bias can be expected to increase with the number of trees sampled or used for calibration.

Table 3. Estimates of the fixed parameters and variance components for the random effects and approximate significance tests for basic mixed model II (Equation 6).

	Estimate	Approximate SE	t value	Approximate P value	Z value	Approximate Z value
Fixed parameter						
a_1	18.97	0.3751	50.58	<0.001		
a_2	−5.471	0.1752	−31.23	<0.001		
Variance component for the random effects						
$\mathrm{Var}(u_{1i})$	26.28	2.842			9.25	<0.001
$\mathrm{Var}(u_{2i})$	4.687	0.5841			8.02	<0.001
$\mathrm{Cov}(u_{1i}, u_{2i})$	−8.406	1.174			−7.16	<0.001
σ^2 (error variance)	2.493	0.03886			64.17	<0.001

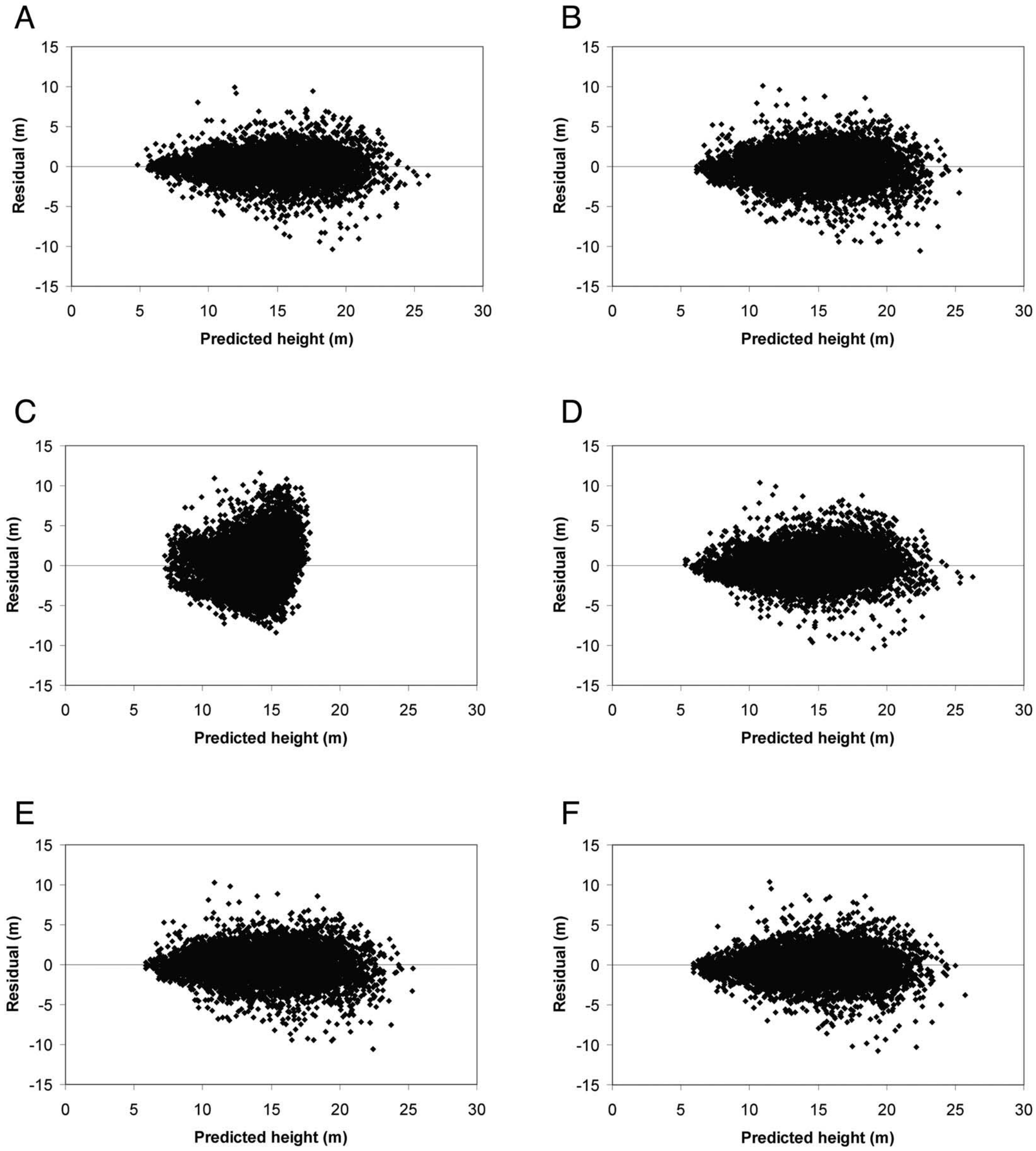

Figure 1. Plots of residuals against predicted total height for (A) the local fit of the basic fixed model I, fitted by ONLS for each plot-inventory combination with all observations, (B) the generalized fixed model III (Equation 7), (C) the mean response (considering only fixed parameters) of the basic mixed model II (Equation 6), (D) Equation 6 calibrated with the total height of the 3 trees which are the diameter quartiles (*Q*), (E) the mean response of the generalized mixed model IV (Equation 8), and (F) Equation 8 calibrated with the total height of the 3 smallest diameter trees.

Results

Table 2 shows the RMSE values of the basic fixed model I fitted by ONLS with all available observations and with 5–15 randomly selected trees for each plot-inventory combination. In the latter case, the RMSE values decreased sharply as the number of sample trees increased from 5 to 6 (a 43.6% reduction in RMSE) and from 6 to 7 (10.4% reduction), and thereafter the RMSE decreased more gradually (see also Figure 2).

Table 4. **Basic mixed model II (Equation 6): RMSE values for the subject-specific, mean (considering only fixed parameters) and calibrated responses obtained with different tree selection strategies.**

Model	RMSE (m)
Subject-specific response	1.543
Mean response	2.905
Calibrated response with	
1 random tree	2.200 (1.629–4.202)
2 random trees	1.967 (1.571–3.661)
3 random trees	1.861 (1.562–3.159)
4 random trees	1.800 (1.556–2.810)
5 random trees	1.761 (1.553–2.564)
6 random trees	1.735 (1.552–2.416)
7 random trees	1.715 (1.551–2.335)
8 random trees	1.699 (1.551–2.261)
9 random trees	1.688 (1.550–2.212)
10 random trees	1.676 (1.550–2.132)
11 random trees	1.670 (1.551–2.090)
12 random trees	1.663 (1.551–2.061)
13 random trees	1.656 (1.550–2.020)
14 random trees	1.651 (1.551–2.007)
15 random trees	1.646 (1.551–1.977)
n-H_i	1.815
3 smallest diameter trees	2.541
Q	1.840
K	1.807

The mean and minimum and maximum (in parentheses) RMSE values for the total height of 1−15 randomly selected trees after 100 simulations are shown along with the total height of the trees used to calculate the dominant height (n-H_i), total height of the 3 trees with smallest diameter, total height of the 3 trees that are the diameter quartiles (Q), and total height of the 4 trees that are the diameter quintiles (K).

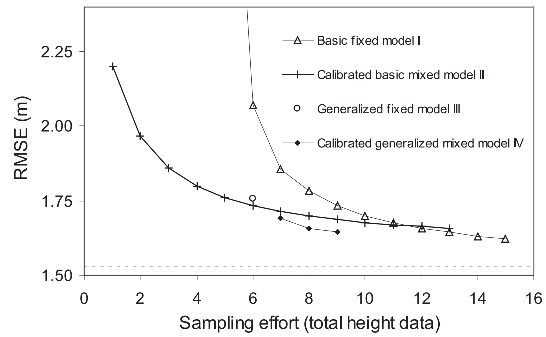

Figure 2. Changes in RMSE with sampling effort for the basic fixed model I fitted locally by ONLS with 5–15 randomly selected trees per plot inventory-combination (△), the basic mixed model II (Equation 6) calibrated with data from 1–15 randomly selected trees (+), the generalized fixed model III (Equation 7) (○), and the generalized mixed model IV (Equation 8) calibrated with data from 1–3 smallest diameter trees (◇). The lower horizontal dashed line represents the RMSE obtained in a local fit of the basic fixed model I, fitted by ONLS for each plot-inventory combination with all observations.

Table 5. Parameter estimates and approximate significance tests of the generalized fixed model III fitted by ONLS (Equation 7).

Parameter	Estimate	Approximate SE	t value	Approximate P value
a_1	2.461	0.2552	9.64	<0.001
a_2	−0.2336	0.0180	−12.99	<0.001
a_3	−0.2737	0.0167	−16.40	<0.001

The estimates of the fixed parameters and variance components for the random effects in the basic mixed model II, all of which were significant at the 0.1% level, are shown in Table 3. The fitted equation is expressed as follows

$$\hat{h}_{ij} = 1.3 + (18.97 + u_{1i})\exp\left(\frac{-5.471 + u_{2i}}{d_{ij}}\right) \quad (6)$$

where $\hat{h}_{ij}$ is the estimated total height (m) of the jth tree in the ith plot-inventory combination, and d_{ij} is its dbh (cm).

Plotting the residuals against estimated total heights for Equation 6 revealed a random pattern of residuals around zero, with homogeneous variance and no detectable significant trends for the subject-specific response, but not for the mean response (considering only fixed parameters, Figure 1C). The RMSE values for the subject-specific, mean, and calibrated responses obtained using different tree selection strategies are shown in Table 4. Calibration with total height of the trees used to estimate the dominant height (with an average of 5.7 trees per plot-inventory combination in this study) yielded an error that was comparable to the average error obtained in the calibration including 4 randomly selected trees and that was slightly greater than the error of the generalized fixed model III. Calibration with total height of randomly selected trees yielded RMSE values that decreased sharply as the number of trees used for the calibration increased, until 3 trees were selected (Figure 2).

The goodness-of-fit statistics of the generalized fixed-model III fitted by ONLS were as follows: EF = 0.776 and RMSE = 1.757 m. All parameters were significant at the 0.1% level (Table 5), and the plot of residuals against estimated total heights showed a random pattern of residuals around zero with homogeneous variance and no detectable significant trends (Figure 1B). The fitted equation is expressed as follows

$$\hat{h}_{ij} = 1.3 + (H_i - 1.3) \times \exp\left((2.461 - 0.2336\,H_i - 0.2737\,d_{gi})\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right) \quad (7)$$

where d_{gi}, H_i, and d_{0i}, are the quadratic mean diameter (cm), dominant height (m), and dominant diameter (cm) of the ith plot-inventory combination, respectively, and the remaining variables were defined previously.

In the generalized mixed model IV, convergence was not achieved when random effects were included in all fixed parameters; therefore, the three possible expansions of two fixed parameters with random effects were evaluated. The estimates of these parameters and of the variance components for the random effects were all significant only at the 5% level when the random effects were included in fixed parameters a_1 and a_3 (Table 6). The fitted equation is expressed as follows

$$\hat{h}_{ij} = 1.3 + (H_i - 1.3)\exp\left((1.771 + u_{1i} - 0.2098\,H_i + (-0.2650 + u_{2i})d_{gi})\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right) \quad (8)$$

The RMSE values for the subject-specific, mean, and calibrated responses obtained using different tree selection strategies are shown in Table 7. Because dominant height is necessary for applying Equation 8, the alternative of calibration with dominant trees was not evaluated.

Table 6. Estimates of the fixed parameters and variance components for the random effects, and approximate significance tests for the generalized mixed model IV (Equation 8).

	Estimate	Approximate SE	*t* value	Approximate *P*-value	*Z* value	Approximate *Z*-value
Fixed parameter						
a_1	1.771	0.6749	2.625	0.009		
a_2	−0.2098	0.05202	−4.033	<0.001		
a_3	−0.2650	0.05421	−4.889	<0.001		
Variance component for the random effects						
$\mathrm{Var}(u_{1i})$	9.571	2.092			4.574	<0.001
$\mathrm{Var}(u_{2i})$	0.06779	0.01022			6.633	<0.001
$\mathrm{Cov}(u_{1i}, u_{2i})$	−0.7298	0.1361			−5.361	<0.001
σ^2 (error variance)	2.496	0.03885			64.26	<0.001

Table 7. Generalized mixed model IV (Equation 8): RMSE values for the subject-specific, mean (considering only fixed parameters), and calibrated responses obtained with different tree selection strategies.

Model	RMSE (m)
Subject-specific response	1.566
Mean response	1.759
Calibrated response with	
1 random tree	1.729 (1.580–2.170)
2 random trees	1.706 (1.572–2.206)
3 random trees	1.687 (1.569–2.174)
4 random trees	1.673 (1.567–2.143)
5 random trees	1.658 (1.566–2.108)
6 random trees	1.647 (1.565–2.057)
7 random trees	1.637 (1.564–2.011)
8 random trees	1.630 (1.564–1.980)
9 random trees	1.622 (1.564–1.931)
10 random trees	1.616 (1.563–1.893)
1 smallest diameter	1.691
2 smallest diameters	1.658
3 smallest diameters	1.645
Q	1.702
K	1.695

The mean and minimum and maximum (in parentheses) RMSE values for total height of 1–10 randomly selected trees after 100 simulations are shown, along with the total height of the 1–3 trees with smallest diameter, total height of the 3 trees that are the diameter quartiles (*Q*), and total height of the 4 trees that are the diameter quintiles (*K*).

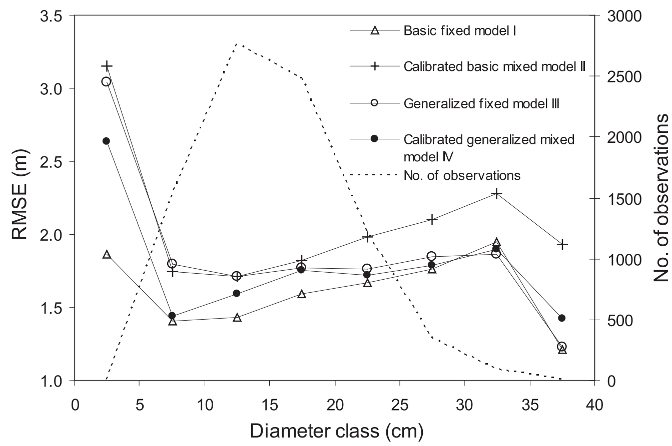

Figure 3. Plot of RMSE against diameter classes for the different model forms evaluated in the study: the basic fixed model I locally fitted by ONLS with data from all trees (△), the basic mixed model II (Equation 6) calibrated with the diameter quartiles (+), the generalized fixed model III (Equation 7) (○), and the generalized mixed model IV (Equation 8) calibrated with data from the 3 smallest diameter trees (●).

Discussion and Conclusions

In the present study, the following models and methodologies were tested and compared to develop a *h-d* relationship: a basic model fitted locally by ONLS for each plot-inventory combination (model I), a basic mixed model (model II), a generalized model fitted by ONLS (model III), and a generalized mixed model (model IV). The RMSE of model I locally fitted and averaged across the complete data set was 1.532 m, which provides a reference of minimum error for comparison with the other model forms. The same model fitted over all data (RMSE = 2.804 m) provides a reference for the maximum error allowable.

The generalized fixed model III included dominant diameter and dominant height in its formulation. Models that include dominant height have an advantage over those including mean height because measurement of dominant height involves less sampling effort (López Sánchez et al. 2003). Moreover, dominant height is usually measured to calculate site index or is included as a regressor variable in several submodels of forest growth simulators. The model also includes a measure of stand density, i.e., quadratic mean diameter. Stand density is the most obvious factor affecting the *h-d* relationship in a stand (Zhang et al. 1997, Zeide and Vanderschaaf 2002, Russell et al. 2010): for a given height, trees in dense stands typically have smaller diameters than those in less dense stands.

If the heights of dominant trees are the only data available, it is preferable to use the generalized fixed model (Equation 7, RMSE = 1.757 m) rather than the basic mixed model (Equation 6) calibrated with dominant tree heights (average RMSE = 1.815 m). The results suggest that the use of the calibrated basic mixed model is an advisable option for tree height estimation when the dominant height is not known. The problem associated with the use of randomly selected trees in a locally fitted basic fixed model or in calibration of a basic or a generalized mixed model is that the error is expressed as an average value, although it actually varies over an interval (see Tables 2, 4, and 7), within which the maximum value is much higher than the mean value.

When a calibrated basic mixed model is used rather than a basic fixed model fitted with data from randomly selected trees, the improvements are more evident when height data are scarce. However, if the number of randomly selected trees is more than 11 (approximately 25% of the trees for an average plot), a locally fitted basic fixed model will provide more accurate results than a calibrated basic mixed model (see Figure 2). Calibration of a basic mixed model with data from one randomly selected tree (as done by Trincado et al. 2007) did not perform adequately in this study (mean RMSE = 2.2 m, with a maximum RMSE = 4.202 m). The use of the total

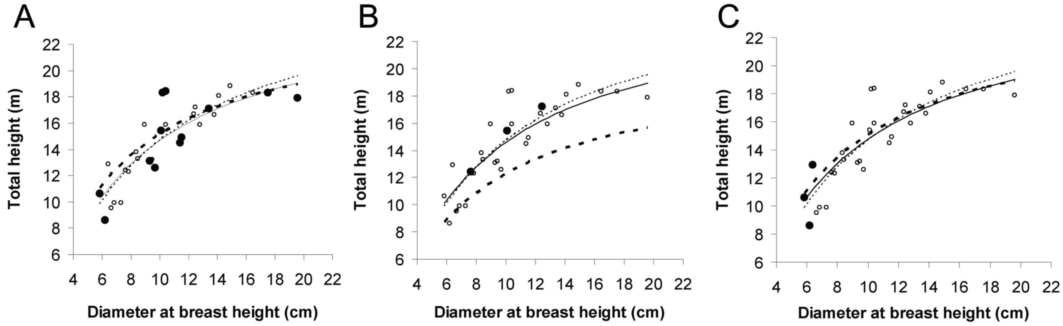

Figure 4. Example of height estimation and comparison with the basic fixed model I locally fitted by ONLS with data from all trees (thin dashed line). This example comprises (A) height estimation from the generalized fixed model III (dashed line) and local fitting (solid line) by the basic fixed model I with data from 12 randomly selected trees (solid points); (B) height estimation by the basic mixed model II for the mean (dashed line) and calibrated (solid line) responses (the calibration was done with the quartiles of the diameter distribution, solid points); and (C) height estimation by the generalized mixed model IV for the mean (dashed line) and calibrated (solid line) responses (calibration was done with the 3 smallest diameters trees, solid points).

height of the 3 trees that are the diameter quartiles (*Q*) presented an adequate balance between the error (RMSE = 1.840 m) and the sampling effort.

All the methodologies analyzed provided similar and unbiased distributions of the residuals against the predicted values (Figure 1) and, in general, there were no problems related to heteroscedasticity. The only exception is the mean response of the basic mixed model, which exhibited heteroscedasticity and was constricted in its predictions; therefore, the basic mixed model should always be calibrated.

The error of the mean response from the generalized mixed model (fixed parameters only, RMSE = 1.759 m) was similar to that obtained from the generalized fixed model (Equation 7, RMSE = 1.757 m). Therefore, when data are not available for calibrating the generalized mixed model, the use of more complex fitting techniques seems unnecessary because the benefits are not obvious. Inventory plots of radius 10 m are very common in forest inventories in Spain, so that the height of the 3–4 largest diameter trees per plot must be measured to enable calculation of the dominant height. However, in the present study, this calibration option does not make sense because dominant height is already considered as a fixed effect in the generalized mixed model, constraining the model by the dominant height-dominant diameter condition (Castedo Dorado et al. 2006). In this study, measurement of the smallest diameter trees improved the accuracy over that of the fixed model and even over that of the model calibrated with randomly selected trees. Calibration with the smallest trees was also considered adequate by Castedo Dorado et al. (2006) and by Crecente-Campo et al. (2010), who used generalized mixed models that included dominant height and dominant diameter as stand-predictor variables, in a way similar to that in the present study.

The total height of the trees used to calculate dominant height must be known to enable use of the generalized fixed model (Equation 7). In the present study, this implied the use of approximately 6 heights per plot-inventory combination or, in relative terms, 13% of all heights. After 100 simulations, the total height of 9 trees per plot-inventory combination (21% of the trees) were required to achieve an overall smaller error in the basic fixed model local fit, whereas for the basic mixed model it was necessary to use 6 randomly selected trees for calibration. Therefore, the use of the basic mixed model calibrated with 6 randomly selected trees involved a sampling effort and an RMSE similar to those for the generalized fixed model (RMSE = 1.757) for the average plot size used in this study. Moreover, the basic mixed model allows a simple model structure to be maintained (Trincado et al. 2007). Nevertheless, one advantage of the generalized fixed model is that it can be used in the disaggregation system of the dynamic stand growth model developed for the species in the region, which includes a dominant height over time prediction function (Gómez-García 2011).

All of the accuracies reported are averaged across all trees. To analyze the behavior of the four model forms evaluated for different tree sizes, the RMSE was calculated and plotted against diameter classes of width 5 cm (Figure 3). Two trends became evident when the first and last diameter classes (0–5 and 35–40 cm) were excluded because they included very few observations. For the basic model (fitted by both ONLS and mixed-model techniques) and for the generalized mixed model, the RMSE values increased with tree size. Nevertheless, for the generalized fixed model, the RMSE values remained quite constant for all diameter classes, which is an additional advantage over the other approaches.

In summary, on the basis of the RMSE values obtained (Figure 2), if a sample of heights of randomly selected trees is measured, use of the calibrated basic mixed model (if the number of tree heights is between 4 and 11) or the locally fitted basic fixed model (if the number of tree heights is greater than 11) is recommended. If only a small number of tree heights are measured, to reduce the sampling effort involved in measuring heights, the basic mixed model should be calibrated with the 3 trees that are the diameter quartiles. The generalized fixed model can be recommended only where dominant height alone is measured. Finally, if information about dominant height and additional heights of the smaller diameter trees is available, the calibrated generalized mixed model could be used. The locally fitted basic fixed model I has the advantage that it is more flexible for application to new stand types or site conditions and does not require prior investment in data collection and model fitting. Although the remaining model forms (which require less sampling effort for tree height prediction) have been developed from a large and widely distributed set of data for the target region, they are subject to regional ecological variations (e.g., related to climate) that could affect their accuracy in future applications. Model development should therefore continue as additional data are collected.

Literature Cited

BURKHART, H.E., AND M.R. STRUB. 1974. A model for simulation of planted loblolly pine stands. P. 128–135 in *Growth models for tree and stand simulation.* Royal College of Forestry, Stockholm, Sweden.

CALAMA, R., AND G. MONTERO. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Can. J. For. Res.* 34:150–163.

CASTEDO DORADO, F., U. DIÉGUEZ-ARANDA, M. BARRIO ANTA, M. SANCHEZ RODRÍGUEZ, AND K.V. GADOW. 2006. A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. *For. Ecol. Manage.* 229:202–213.

CASTROVIEJO, S., M. LAÍNZ, G. LÓPEZ GONZÁLEZ, P. MONTESERRAT, F. MUÑOZ GARMENDIA, J. PAIVA, AND L. VILLAR. 1990. *Flora ibérica: Plantas vasculares de la Península Ibérica e Islas Baleares. Volumen II: Platanaceae-Plumbaginaceae* (partim.). Real Jardín Botánico (CSIC), Madrid, Spain. 897 p.

CRECENTE-CAMPO, F., M. TOMÉ, P. SOARES, AND U. DIÉGUEZ-ARANDA. 2010. A generalized nonlinear mixed effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. *For. Ecol. Manage.* 259:943–952.

CURTIS, R.O. 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *For. Sci.* 13(4):365–375.

FANG, Z., AND R.L. BAILEY. 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For. Sci.* 47:287–300.

GÓMEZ-GARCÍA, E. 2011. *Modelos dinámicos de crecimiento para rodales regulares de Betula pubescens Ehrh. y Quercus robur L. en Galicia.* PhD thesis, Universidade de Santiago de Compostela, A Coruña, Spain. 157 p.

HUANG, S., S.X. MENG, AND Y. YANG. 2009. Assessing the goodness of fit of forest models estimated by nonlinear mixed-models methods. *Can. J. For. Res.* 39:2418–2436.

KRUMLAND, B.E., AND L.C. WENSEL. 1988. A generalized height-diameter equation for coastal California species. *West. J. Appl. For.* 3:113–115.

LAPPI, J. 1997. A longitudinal analysis of height/diameter curves. *For. Sci.* 43:555–570.

LARSEN, D.R., AND D.W. HANN. 1987. *Height-diameter equations for seventeen tree species in Southwest Oregon.* Forest Research Laboratory, Res. Paper 49, Oregon State University, Corvallis, OR. 16 p.

LINDSTROM, M.J., AND D.M. BATES. 1990. Nonlinear mixed-effects models for repeated measures data. *Biometrics* 46:673–687.

LÓPEZ SÁNCHEZ, C.A., J.J. GORGOSO, F. CASTEDO DORADO, A. ROJO, R. RODRÍGUEZ, J.G. ÁLVAREZ GONZÁLEZ, AND F. SÁNCHEZ. 2003. A height-diameter model for *Pinus radiata D.* Don in Galicia (Northwest Spain). *Ann. For. Sci.* 60:237–245.

MEHTÄTALO, L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. *Can. J. For. Res.* 34(1):131–140.

PEÑA, D. 2002. *Regresión y diseño de experimentos*. Alianza Editorial, Madrid, Spain. 744 p.

PENG, C. 2001. Developing and validating nonlinear height-diameter models for major tree species of Ontario's boreal forest. *North. J. Appl. For.* 18:87–94.

PINHEIRO, J.C., AND D.M. BATES. 1995. *Model building for nonlinear mixed effects model.* Department of Biostatistics, Tech. Rep. 91, University of Wisconsin-Madison, Madison, WI. 11 p.

PRODAN, M., R. PETERS, F. COX, AND P. REAL. 1997. *Mensura forestal.* Instituto Interamericano de Cooperación para la Agricultura (IICA)/Deutsche Gesellschaft für Technische Zusammernarbeit (GTZ), San José, Costa Rica. 561 p.

RUSSELL, M.B., R.L. AMATEIS, AND H.E. BURKHART. 2010. Implementing regional locale and thinning response in the loblolly pine height-diameter relationship. *South. J. Appl. For.* 34(1):21–27.

SAS INSTITUTE, INC. 2009. *SAS/STAT 9.2 user's guide*, 2nd ed. SAS Institute Inc., Cary, NC. 7869 p.

SHARMA, M., AND J. PARTON. 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *For. Ecol. Manage.* 249:187–198.

SHARMA, M., AND S.Y. ZHANG. 2004. Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana. Scand. J. For. Res.* 19:442–451.

TEMESGEN, H., AND K.V. GADOW. 2004. Generalized height-diameter models-an application for major tree species in complex stands of interior British Columbia. *Eur. J. For. Res.* 123:45–51.

TOMÉ, M. 1988. *Modelação do crescimento da árvore individual em povoamentos de* Eucalyptus globulus *Labill.* (*1ª rotação*) *na região centro de Portugal.* PhD thesis, Instituto Superior de Agronomía, Universidade Técnica de Lisboa, Lisbon, Portugal. 256 p.

TRINCADO, G., L.V. CURTIS, AND H.E. BURKHART. 2007. Regional mixed-effects height- diameter models for loblolly pine (*Pinus taeda L.*) plantations. *Eur. J. For. Res.* 126:253–262.

VONESH, E.F., AND V.M. CHINCHILLI. 1997. *Linear and nonlinear models for the analysis of repeated measurements.* Marcel Dekker Inc., New York. 560 p.

XUNTA DE GALICIA. 2001. *O monte galego en cifras.* Dirección Xeral de Montes e Medio Ambiente Natural, Consellería de Medio Ambiente, Santiago de Compostela, Spain. 226 p.

ZEIDE, B., AND C. VANDERSCHAAF. 2002. The effect of density on the height-diameter relationship. P. 463–466 in *Proc. of the 11th Biennial southern silvicultural research conference, 2001 March 20–22, Knoxville, TN*, Outcalt, K.W. (ed.). USDA For. Serv., Gen. Tech. Rep. SRS-48, Asheville, NC.

ZHANG, S.A., H.E. BURKHART, AND R.L. AMATEIS. 1997. The influence of thinning on tree height and diameter relationships in loblolly pine plantations. *South. J. Appl. For.* 21:199–205.

Appendix: Height Estimation—An Example

In a new plot, the dbh of every tree is measured and the quadratic mean diameter (d_{gi}) and dominant diameter (d_{0i}) can therefore be calculated. Assuming an example in which $d_{gi} = 10.80$ cm and $d_{0i} = 17.12$ cm and also the total height of some trees is known, we will consider four cases: (1) total height of the trees corresponding to the quartiles of the diameter distribution (*h-d* pairs of 12.4 m−7.65 cm, 15.4 m−10.10 cm, and 17.2 m−12.45 cm); (2) mean height of dominant trees ($H_i = 18.3$ m); (3) mean height of dominant trees plus total height of the 3 smallest diameter trees (*h-d* pairs of 10.6 m−5.85 cm, 8.6 m−6.2 cm, and 12.9 m−6.4 cm), and (4) total height of 12 randomly selected trees. The example aims to describe the calibration of both the basic (Equation 6) and generalized (Equation 8) mixed models. Figure 4 shows the *h-d* curves obtained for these calibrations, including those obtained from the generalized fixed model (Equation 7) and from the basic fixed model I that was fitted locally with the data corresponding to the randomly selected trees.

Basic Mixed Model

The basic mixed model will be calibrated with the trees corresponding to the quartiles of the diameter distribution. The estimated variances and the covariance of the random effects (Table 3) are the elements of the variance-covariance matrix $\hat{\mathbf{D}}$

$$\hat{\mathbf{D}} = \begin{bmatrix} 26.28 & -8.406 \\ -8.406 & 4.687 \end{bmatrix}$$

The variance-covariance matrix for the random error term is determined by assuming that all estimations have constant variance ($\hat{\sigma}^2$) (Table 3) and that the errors are not correlated

$$\hat{\mathbf{R}}_k = \hat{\sigma}^2 \times \mathbf{I}_3 = \begin{bmatrix} 2.493 & 0 & 0 \\ 0 & 2.493 & 0 \\ 0 & 0 & 2.493 \end{bmatrix}$$

where $\mathbf{I}_3$ is the identity matrix with dimension (3×3) equal to the number of data used for calibration.

The partial derivatives with respect to the random effects u_{1i} and u_{2i} are $Z_{u_1} = \exp(a_2/d_{ij})$ and $Z_{u_2} = a_1 \exp(a_2/d_{ij})(1/d_{ij})$. Therefore, the partial derivatives matrix with respect to the random effects $\mathbf{Z}_k$ is

$$\mathbf{Z}_k = \begin{bmatrix} 0.4891 & 1.213 \\ 0.5818 & 1.093 \\ 0.6444 & 0.9820 \end{bmatrix}$$

The matrix $\hat{\mathbf{e}}_k$ of residuals of total height obtained with the basic mixed model mean response (only fixed parameters; Equation 6 with u_1 and u_2 equal 0) is

$$\hat{\mathbf{e}}_k = \begin{bmatrix} 1.823 \\ 3.063 \\ 3.675 \end{bmatrix}$$

Therefore, the random effects predicted with Equation 3 were $\hat{\mathbf{u}}_k = [4.733 - 0.3150]^T$. Thus, using Equation 6, the calibrated response is expressed as follows

$$\hat{h}_{ij} = 1.3 + (18.97 + 4.733)\exp\left(\frac{-5.471 - 0.3150}{d_{ij}}\right)$$

The mean and calibrated responses (calibration with quartiles) of the basic mixed model used in this example are shown in Figure 4B, which also shows a local fit of the basic fixed model I with the complete data set.

Generalized Mixed Model

In a way similar to that in the previous example, the generalized mixed model will be calibrated with the 3 smallest diameter trees. The estimated variances and the covariance of the random effects (Table 6) are the elements of the variance-covariance matrix $\hat{\mathbf{D}}$

$$\hat{\mathbf{D}} = \begin{bmatrix} 9.571 & -0.7298 \\ -0.7298 & 0.06779 \end{bmatrix}$$

The variance-covariance matrix for the random error term is determined by assuming that all estimations have constant variance ($\hat{\sigma}^2$) (Table 6) and that the errors are not correlated

$$\hat{\mathbf{R}}_k = \hat{\sigma}^2 \times \mathbf{I}_3 = \begin{bmatrix} 2.496 & 0 & 0 \\ 0 & 2.496 & 0 \\ 0 & 0 & 2.496 \end{bmatrix}$$

The partial derivatives with respect to the random effects u_{1i} and u_{2i} are $Z_{u_1} = (H_i - 1.3)\exp\left[(a_1 + a_2 H_i + a_3 d_{gi}) \times \left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right]\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)$ and $Z_{u_2} = (H_i - 1.3)\exp\left[(a_1 + a_2 H_i + a_3 d_{gi})\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right]\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)d_{gi}$. Therefore, the partial derivatives matrix with respect to the random effects $\mathbf{Z}_k$ is

$$\mathbf{Z}_k = \begin{bmatrix} 1.099 & 11.88 \\ 1.054 & 11.39 \\ 1.028 & 11.10 \end{bmatrix}$$

The matrix $\hat{\mathbf{e}}_k$ of residuals of total height obtained with the generalized mixed model mean response (only fixed parameters; Equation 8 with u_1 and u_2 equal 0) is

$$\hat{\mathbf{e}}_k = \begin{bmatrix} -0.4672 \\ 2.944 \\ 1.098 \end{bmatrix}$$

Therefore, the random effects predicted with Equation 3) were $\hat{\mathbf{u}}_k = [0.5061 - 0.0008174]^T$. Thus, using Equation 8, the calibrated response is expressed as follows

$$\hat{h}_{ij} = 1.3 + (H_i - 1.3) \times \exp\left((1.771 - 0.5061 - 0.2098\, H_i + (-0.2650 - 0.0008174)d_{gi})\left(\frac{1}{d_{ij}} - \frac{1}{d_{0i}}\right)\right)$$

The mean and calibrated responses (calibration with smallest diameter trees) of the generalized mixed model used in this example are shown in Figure 4C, which also shows a local fit of the basic fixed model I with the complete data set.