

# Selection of mixed-effects parameters in a variable–exponent taper equation for birch trees in northwestern Spain

Esteban Gómez-García · Felipe Crecente-Campo ·  
Ulises Diéguez-Aranda

Received: 1 April 2013 / Accepted: 1 July 2013 / Published online: 31 July 2013  
© INRA and Springer-Verlag France 2013

## Abstract

• **Context** Taper equations predict the variation in diameter along the stem, therefore characterizing stem form. Several recent studies have tested mixed models for developing taper equations. Mixed-effects modeling allow the interindividual variation to be explained by considering both fixed-effects parameters (common to the population) and random-effects parameters (specific to each individual).  
• **Aims** The objective of this study is to develop a mixed-effect variable exponent taper equation for birch trees in northwestern Spain by determining which fixed-effects parameters should be expanded with random-effects parameters.  
• **Methods** All possible combinations of linear expansions with random effects in one and in two of the fixed-effects model parameters were tested. Upper stem diameter measurements were used to estimate random-effects parameters by the use of an approximate Bayesian estimator, which calibrated stem profile curves for individual trees.  
• **Results** Parameter estimates for more than half of the mixed models investigated were nonsignificant. A first order autoregressive error structure was used to completely remove

the autocorrelation between residuals, as mixed-effects modeling were not sufficient for this purpose.

• **Conclusion** The mixed model with the best fitting statistics did not provide the best calibration statistics for all upper stem diameter measurements. From a practical point of view, model calibration should be considered an essential criterion in mixed model selection.

**Keywords** *Betula pubescens* Ehrh · Taper equation · Mixed-effects modeling · Autocorrelation · Calibration · Galicia

## 1 Introduction

Taper equations predict the variation in diameter along the stem, and they therefore characterize stem form (Clutter et al. 1983; Avery and Burkhart 2002; Kozak 2004). They are especially useful when product demand is expressed in terms of tree dimensions or assortments and allows for predictions of product mixes and hence the economic value of the tree (Fonweban et al. 2011). Taper equations provide estimates of the following variables (Kozak 2004): (1) diameter at any height along the stem, (2) total stem volume, (3) merchantable volume and height to any top diameter and from any stump height, and (4) volumes of logs of any length at any height from the ground.

Taper functions are usually classified as single, segmented, or variable–exponent; the latter of these usually involve less bias and greater precision in estimating diameters at different heights (e.g., Kozak 1988, 2004; Pérez et al. 1990; Newnham 1992; Muhairwe 1999). Assuming that stem form varies continuously with stem height, variable–exponent models describe stem shape with a changing exponent or variable from ground to top to represent the neiloid, paraboloid, conic, and several intermediate forms (Kozak 1988; Newnham 1988).

---

**Handling Editor:** Barry Alan Gardiner

**Contribution of the co-authors** Esteban Gómez-García and Ulises Diéguez-Aranda analyzed the data and wrote the manuscript. Felipe Crecente-Campo provided technical assistance in model fitting and revised the text.

---

E. Gómez-García (✉) · F. Crecente-Campo · U. Diéguez-Aranda  
Departamento de Ingeniería Agroforestal, Universidad de Santiago  
de Compostela. Escuela Politécnica Superior, R/Benigno Ledo,  
Campus Universitario, 27002 Lugo, Spain  
e-mail: esteban.gomez@usc.es

F. Crecente-Campo  
e-mail: felipe.crecente@usc.es

U. Diéguez-Aranda  
e-mail: ulises.dieguez@usc.es

Since the end of the twentieth century, several studies have used mixed models to develop taper equations (e.g., Tasissa and Burkhart 1998; Valentine and Gregoire 2001; Garber and Maguire 2003; Leites and Robinson 2004; Trincado and Burkhart 2006; Yang et al. 2009; Fonweban et al. 2011). Mixed models allow for both mean and subject-specific responses (Davidian and Giltinan 1995; Littell et al. 2006). The mean response considers only fixed-effects parameters, common to the population, while the subject-specific response considers both fixed- and random-effects parameters common to each subject. Mixed models have been used to deal with correlated errors associated with longitudinal data (Lindstrom and Bates 1990), such as successive measurements of diameter along the stem. An additional advantage of these models is that they can use prior information about a new individual (one or more complementary diameter measurement) to estimate the random-effects parameters, therefore providing a calibrated response. This technique enables calibration of taper equations for trees growing in sites of different quality and under different silvicultural regimes (Trincado and Burkhart 2006).

Garber and Maguire (2003) used mixed models to fit variable–exponent taper equations for three species in Oregon (USA), although they did not demonstrate the utility of using other diameters for calibration. Trincado and Burkhart (2006) used mixed models to fit a segmented taper equation and evaluated the use of one and two upper stem diameter measurements to obtain the random-effects parameters; however, they did not analyze the calibration performance of different random-effects expansion options. In both of these studies, mixed models were unable to completely eliminate the autocorrelation, for which they used continuous autoregressive error structures. Yang et al. (2009) used mixed models to fit variable–exponent taper equations but did not fully correct the autocorrelation and evaluated model calibration in a similar way as Trincado and Burkhart (2006), although for total and log volume prediction.

Birch (*Betula pubescens* Ehrh., also referred to as *Betula alba* L. and *B. pubescens* subsp. *celtibérica* Rothm. & Vasc., Castroviejo et al. 1990) is one of the main forest species in Galicia (NW Spain). In this region, the species grows between 0 and 1,700 m above sea level, although it is more abundant in the northeastern areas at altitudes above 400–500 m; in the south of the region, it only appears naturally near streams or in wet locations. The species is considered a fast-growing pioneer species, which rapidly colonizes open ground derived from human activity (burning, cutting, or grazing) or natural disturbance. At the beginning of this century, 32,000 ha of forest stands included birch as the main tree species in Galicia (Xunta de Galicia 2001).

The objective of the present study was to develop a mixed-effects variable–exponent taper equation for birch trees in Galicia, based on the Kozak (2004) taper model, by

investigating which fixed-effects parameters should be expanded with random-effects parameters. For this purpose, fitting statistics and practical use of the model by calibration based on an upper stem diameter measurement were considered. The error structure of the data due to residual correlation was also considered.

## 2 Materials and methods

### 2.1 Data

A total of 304 birch trees in 130 research plots were destructively sampled as part of a region-wide study carried out in Galicia by the Sustainable Forest Management Unit of the University of Santiago de Compostela. Diameter at breast height ( $d$  in centimeters, 1.3 m above ground) was measured to the nearest 0.1 cm in each of the trees. The trees were later felled, leaving stumps of average height 0.16 m, and total bole length was measured to the nearest 0.01 m to calculate the total tree height ( $h$ , in meters). The bole was then cut into logs at 1-m intervals until reaching a diameter of 7 cm. Two perpendicular diameters over bark were measured in each cross section (at height  $h_i$  in meters from ground level), to the nearest 0.1 cm, and were then averaged ( $d_i$ , in centimeters).

The scatter plot of relative diameter ( $d_i/d$ ) against relative height ( $h_i/h$ ) was visually examined to detect any possible anomalies in the data. As the data included extreme data points, a systematic procedure for detecting abnormal data points (Bi 2000) was applied to increase the efficiency of the process. This involved local quadratic nonparametric fitting (LOESS) with a smoothing factor of 0.3, which was selected after iterative fitting and visual examination of the smoothed taper curves overlaid on the data. This approach removed 1.04 % of the initial data. Some of the outliers may represent mistakes in measuring the bole sections or in transcription of field notes, but some others may represent deformations brought about by fire damage, large knots, and other types of physical damage such as partial death, growth deformations, etc. These data points were excluded from further analysis. Summary statistics of the data finally used are shown in Table 1.

### 2.2 Models analyzed

Firstly, the models developed by Max and Burkhart (1976), Bi (2000), Fang et al. (2000), and Kozak (2004) (with the modification proposed by Yang et al. 2009) were fitted by ordinary nonlinear least squares (ONLS) in order to evaluate their performance for birch in Galicia. The variable–exponent model of Kozak (2004) was the most accurate for predicting diameters at different heights (based on fitting statistics), and it was therefore chosen for further analysis.

**Table 1** Descriptive statistics of the sample of trees used

Variable	Mean	Minimum	Maximum	Std. dev.
N° sections	10.3	4	21	3.4
$d$	20.0	7.3	39.2	6.0
$h$	14.5	6.2	24.4	3.4
$h_{st}$	0.16	0.0	0.5	0.08

$d$  diameter at breast height (cm),  $h$ , total tree height (m),  $h_{st}$ , stump height (m)

This model has been used successfully in numerous studies and has shown to be highly flexible and can be adapted to different trees and species (e.g., Diéguez-Aranda et al. 2006; Corral-Rivas et al. 2007; Yang et al. 2009).

Yang et al. (2009) used mixed models to analyze four variable–exponent taper equations and concluded that Eq. 1 performed best for lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.) in Alberta. This equation is a modification of the original model of Kozak (2004) by adding parameter  $b_0$  to the variable–exponent:

$$d_i = a_0 d^{a_1} h^{a_2} x^{b_0 + b_1 q^4 + b_2 (1/\exp(d/h)) + b_3 x^{0.1} + b_4 (1/d) + b_5 h^w + b_6 x} \quad (1)$$

where

$$x = \frac{w}{1 - (1.3/h)^{1/3}}$$

$$w = 1 - q^{1/3}$$

$$q = h_i/h$$

$a_0$ – $a_2$  and  $b_0$ – $b_6$  are fixed-effects parameters to be estimated.

### 2.3 Mixed-effects modeling

A mixed model includes both fixed (common to the population) and random (subject-specific) effects. A key question when developing mixed models is which parameters should be considered as fixed and which as mixed (both fixed and random effects). One possible approach is to fit each individual independently, without considering the random effects, and then to expand the most variable parameters by use of random effects (Fang and Bailey 2001). However, this approach requires sufficient observations for each individual for the estimated parameters to be significant in each individual fit, which is not the case in the present study (the fitted fixed-effects model includes ten parameters). Another possible approach is to fit different combinations of fixed- and random-effects parameters and to choose the best fitted mixed model. Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC) are often used to determine the

best combination of fixed and random effects.

$$\text{AIC} = -2\ln(L) + 2\lambda \quad (2)$$

$$\text{BIC} = -2\ln(L) + \lambda \ln(m) \quad (3)$$

Where  $L$  is the maximum likelihood (ML) value;  $\lambda$  is the number of effective parameters, calculated by adding the number of fixed-effects parameters,  $p$ , and the number of variance and covariance parameters estimates; and  $m$  is the number of trees for mixed models and the total number of observations for fixed-effects models.

Although some studies have indicated that the correlations within each tree can be largely overcome by modeling random effects with an unstructured variance–covariance matrix (e.g., Vonesh and Chinchilli 1997; VanderSchaaf and Burkhart 2007; Yang et al. 2009), other studies have shown that this is not always possible (e.g., Garber and Maguire 2003; Trincado and Burkhart 2006). A first order autoregressive covariance structure (AR (1)), because the measurements were made at a constant distance, was used to completely remove autocorrelation between residuals from the same tree. This structure assumes that the errors are correlated, and it models the variance–covariance matrix for the error term ( $\mathbf{R}_k$ ) as follows:  $\sigma^2 \times \mathbf{\Gamma}_k(\rho)$ , where  $\sigma^2$  is the error variance, and  $\mathbf{\Gamma}_k(\rho)$  is a ( $n_k \times n_k$ ) matrix describing the pattern of correlation between the measurements of the individual  $k$ . AR (1) assumes that for the  $ki$ -th element, the covariance structure is  $\rho^{|ki - ki'|}$ .

### 2.4 Parameter estimation

The SAS %NLINMIX macro (SAS support 2011) was used to estimate the fixed-effects parameters and the variance–covariance parameters associated with random-effects parameters. Two methods of expansion can be used in this macro as follows: expansion around zero, which is the expected value for random-effects parameters, or expansion around the empirical best linear unbiased predictor (EBLUP) of the random-effects parameters (Littell et al. 2006). Both approaches produce reliable estimates (Davidian and Giltinan 1993; Pinheiro and Bates 1995; Wolfinger and Lin 1997). Although the EBLUP method sometimes generates slightly better results, it requires more computation time, is less stable, and is very sensitive to model specification (Hartford and Davidian 2000). Moreover, Vonesh (1996) showed that consistent estimates with the EBLUP expansion method could only be obtained if the number of individuals and observations per individual are infinite. For our data, most of the models with various combinations of random-effects parameters failed to converge when the EBLUP expansion method was used. Therefore, all results presented in this study are based on the expansion around zero method.

A ML procedure or a restricted maximum likelihood (REML) procedure can be used to fit the model. The former was used to extract the AIC and BIC statistics for model comparison because the ML values are comparable (unlike in the REML procedure), and the AIC and BIC are based on these values. However, the ML estimates of the variance components do not take into account the degrees of freedom lost in estimating the fixed-effects parameters and are therefore biased downwards (Yang et al. 2009), in contrast to the REML estimates (Laird and Ware 1982; Littell et al. 2006). Therefore, the REML procedure was used to obtain the final parameter estimates of the models, which were then used for testing calibrations with an upper stem diameter measurement.

### 2.5 Calibrated response

Estimation of the random-effects parameters of a new individual (i.e., calibration) requires one or more complementary diameter measurement. This was done by use of an approximate Bayesian estimator (Vonesh and Chinchilli 1997; Trincado and Burkhart 2006):

$$\hat{\mathbf{u}}_k \cong \hat{\mathbf{D}}\mathbf{Z}_k^T (\mathbf{Z}_k \hat{\mathbf{D}}\mathbf{Z}_k^T + \hat{\mathbf{R}}_k)^{-1} \hat{\mathbf{e}}_{ki} \tag{4}$$

Where  $\hat{\mathbf{D}}$  is the estimated variance–covariance matrix for the random-effects parameters  $\mathbf{u}_k$ ,  $\hat{\mathbf{R}}_k$  is the estimated variance–covariance matrix for the error term,  $\hat{\mathbf{e}}_{ki}$  is the error of diameter obtained with the mean response (only fixed-effects parameters), and  $\mathbf{Z}_k$  is the partial derivatives matrix with respect to random-effects parameters:  $\mathbf{Z}_k = \partial f(\mathbf{x}_k, \boldsymbol{\beta}, 0) / \partial \mathbf{u}_k |_{\hat{\boldsymbol{\beta}}, 0}$ . When the random-effects parameters,  $\mathbf{u}_k$ , are introduced linearly in the parameters vector, and when the nonlinear function is linearized by the zero expansion method, the partial derivatives of the  $\mathbf{Z}_k$  matrix are equivalent to the partial derivatives with respect to the fixed-effects parameters (Yang et al. 2009).

A previous analysis showed similar accuracy in calibration with one and two upper stem diameter measurements (as shown by Trincado and Burkhart 2006), and therefore calibration with only one diameter was evaluated in this study. This reduces the cost involved in obtaining upper stem diameter measurements. Similarly to Yang et al. (2009), the following statistics (Cochran 1963) were calculated to evaluate the calibration procedure:

$$\bar{e} = \sum_{i=1}^n (y_i - \hat{y}_i) / n \tag{5}$$

$$\bar{e}\% = 100 \times \bar{e} / \bar{y} \tag{6}$$

$$\text{MSE} = \bar{e}^2 + s^2 \tag{7}$$

Where  $y_i$  and  $\hat{y}_i$  are the diameters outside bark ( $d$ ) observed and calibrated for the  $i$ th observation ( $i=1, 2, \dots, n$ ),  $n$

is the total number of diameters observed,  $\bar{y}$  is the mean of observed diameters,  $\bar{e}$  is the mean prediction error,  $\bar{e}\%$  is the mean prediction error expressed as a percentage of the mean value,  $s^2$  is the variance of the prediction error, and MSE is the mean square error of the predictions. The mean error and the mean percentage prediction error provide an average measure of the prediction bias, with  $\bar{e}$  in absolute and  $\bar{e}\%$  in relative terms. MSE combines the mean bias and the variation in the biases, thus providing a better measure of the accuracy of the model. The MSE was used as the main criterion in evaluating the calibration.

To determine the best calibration height between 2 and 8 m, diameters along the stem were calculated every 0.5 m by linear interpolation between the two nearest stem measurements. The lower limit was selected because it is quite close to breast height. The upper limit was chosen as a compromise between use of the greatest number of trees taller than this threshold and evaluation of as much length as possible along the stem for calibration. Therefore, the number of observations used in calibration was decreased from 3,146 to 2,432 diameter–height pairs. Absolute height, rather than relative height, was evaluated because it is easier to obtain in forestry operations.

### 2.6 Criteria for selection of the final mixed model

All possible expansion options of one and two fixed-effects parameters with random-effects parameters in Eq. 1 were evaluated. Fitting statistics (AIC and BIC) were calculated for the mixed model, and calibration statistics ( $\bar{e}$ ,  $\bar{e}\%$ , and MSE) were calculated for the calibrated mixed model. The final selection of the best alternative was based on practical use of the model, and therefore calibration statistics were the main criteria used in the decision.

## 3 Results

When random-effects parameters were used to expand various fixed-effects parameters in Eq. 1, for 35 out of the 55 possible combinations (10 of one mixed-effects parameter plus 45 of two), some of the fixed-effects parameters or the variance of a random-effects parameter were not significant at a 5 % level. When these parameters were removed and the models refitted, the results were poorer (lower values of fitting statistics) than for the remaining 20 combinations (see Table 2 for AIC and BIC values). Equation 1, in which parameters  $b_0$  and  $b_6$  were linearly expanded with random-effects parameters, was the best mixed model as regards fitting statistics.

The 20 different combinations were then calibrated with one diameter measurement at different heights along the stem. The variations in MSE in diameter estimation with calibration height for these models are shown in Fig. 1. The mixed models were split into two groups that displayed



**Table 2** AIC and BIC values for the mixed models with all fixed-effects parameters and the variance of the random-effects parameters significantly different from zero, at a probability level of 5 % (in ascending order of AIC and BIC)

N°	Expanded fixed-effects parameters	AIC	BIC	N°	Expanded fixed-effects parameters	AIC	BIC
1	$b_0, b_6$	9,339	9,395	11	$a_1, b_2$	9,737	9,793
2	$b_6$	9,506	9,551	12	$a_0, b_2$	9,743	9,798
3	$b_0, b_3$	9,518	9,574	13	$b_1, b_3$	9,843	9,899
4	$a_2, b_3$	9,678	9,734	14	$b_2, b_3$	9,848	9,904
5	$a_1, b_3$	9,681	9,737	15	$b_0, b_1$	9,887	9,943
6	$a_0, b_3$	9,687	9,743	16	$b_3$	9,902	9,951
7	$a_2, b_0$	9,717	9,773	17	$b_1, b_4$	9,903	9,959
8	$a_1, b_0$	9,720	9,776	18	$b_1, b_2$	9,907	9,963
9	$a_0, b_0$	9,726	9,781	19	$b_2, b_4$	9,958	10,010
10	$a_2, b_2$	9,734	9,790	20	$b_2$	9,968	10,016

similar trends. The best calibration results for heights up to 5 m were obtained with the models in group 1. In these models, the slope of line MSE–calibration height was less steep, and the variation in MSE ranged between 2.45 and 2.20 cm<sup>2</sup> for heights of 2 to 8 m. For calibration heights above 5 m, the MSE values were lower for some group 2 models than for group 1 models. However, the calibration performance was less stable, with MSE values ranging from approximately 2.70 to 2.05 cm<sup>2</sup>. From a practical point of view, the greater effort (which increases with height) of measuring an additional diameter for calibration should be taken into account. Thus, the group 1 models are preferable because they are less-dependent on the calibration height and provide better results for calibration heights lower than 5 m. Within group 1, the models that consider the expansion of parameters  $a_0, b_0$ ;  $a_1, b_0$ ; and  $a_2, b_0$  produced the lowest MSE, and within these models, the  $a_1, b_0$  model was slightly superior. Therefore, the following mixed model was finally proposed:

$$d_i = a_0 d^{a_1 + u_{k1}} h^{a_2} x^{b_0 + u_{k2} + b_1 q^4 + b_2 (1/\exp(d/h)) + b_3 x^{0.1} + b_4 (1/d) + b_5 h^w + b_6 x \tag{8}$$

where

$$x = \frac{w}{1 - (1.3/h)^{1/3}}$$

$$w = 1 - q^{1/3}$$

$$q = h_i/h$$

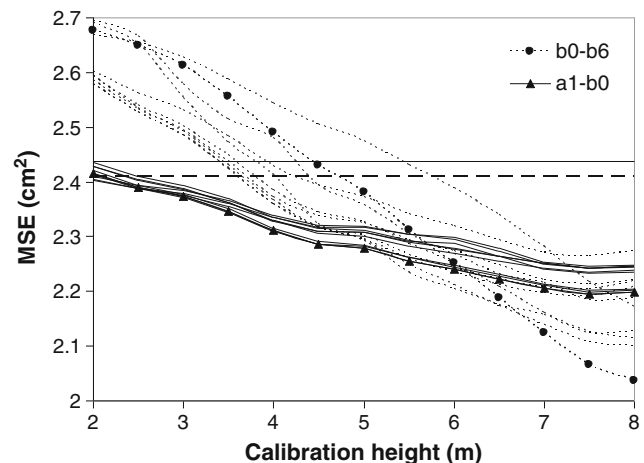
$u_{k1}$  and  $u_{k2}$  are random-effects parameters.

Estimated fixed-effects parameters and the variance components of the random-effects parameters, along with the values of the approximate significance contrasts (all significant at a 5 % level) for Eq. 8, are shown in Table 3. The use of the mixed-effects approach did not completely remove the autocorrelation (Fig. 2b). However, the use of a first order autoregressive error structure AR (1) successfully removed

the autocorrelation (Fig. 2c). The autoregressive parameter ( $\rho$ ) was equal to 0.6656. The mean response (only with fixed-effects parameters) of the selected mixed model (Eq. 8) provided a similar accuracy (MSE=2.44 cm<sup>2</sup>) as Eq. 1 fitted by ONLS and correction for autocorrelation (MSE=2.41 cm<sup>2</sup>), as shown in Fig. 1. The MSE of the mean response of the 20 models shown in Table 2 ranged from 2.43 to 2.55 cm<sup>2</sup>.

### 3.1 Mean response and calibration—an example

To illustrate the calibration process, upper stem diameters for a tree of diameter at breast height  $d=33.3$  cm and total height



**Fig. 1** Variation in MSE of the calibrated mixed models with calibration height for the models shown in Table 2. Two different groups are considered: group 1 (solid lines), expanding the following fixed-effects parameters:  $a_2, b_2$ ;  $a_1, b_2$ ;  $a_0, b_2$ ;  $a_2, b_0$ ;  $a_1, b_0$ ;  $a_0, b_0$ ;  $a_2, b_3$ ;  $a_1, b_3$ ;  $a_0, b_3$ ;  $b_6$ ; and group 2 (dotted lines), expanding the following fixed-effects parameters:  $b_0, b_1$ ;  $b_1, b_3$ ;  $b_1, b_2$ ;  $b_2$ ;  $b_3$ ;  $b_1, b_4$ ;  $b_0, b_3$ ;  $b_0, b_6$ ;  $b_3, b_4$ ; and  $b_2, b_3$ . The MSE of the fixed part of the “ $a_1, b_0$ ” mixed model (i.e., the mean response: horizontal solid line) and the MSE of Eq. 1 fitted with ONLS and correction for autocorrelation (horizontal dashed line) are also shown

**Table 3** Fixed-effects parameter estimates, approximate significance contrasts, and variance components of Eq. 8 fitted by the restricted maximum likelihood procedure (REML)

Parameter	Estimate	Approx. std. error	Approx. <i>t</i> value	Approx. prob. >   <i>t</i>
$a_0$	0.9652	0.0537	17.96	<0.001
$a_1$	0.9421	0.0198	47.53	<0.001
$a_2$	0.08482	0.0266	3.19	0.001
$b_0$	4.094	0.581	7.04	<0.001
$b_1$	-0.3939	0.138	-2.85	0.004
$b_2$	-0.4112	0.129	-3.18	0.001
$b_3$	-4.153	0.660	-6.29	<0.001
$b_4$	2.813	0.628	4.48	<0.001
$b_5$	0.05345	0.00492	10.87	<0.001
$b_6$	0.3332	0.0847	3.93	<0.001
Variance component	Estimate	Approx. std. error	Approx. <i>Z</i> -value	Approx. prob. >   <i>Z</i>
var( $u_{k1}$ )	0.0003373	0.00004	7.50	<0.001
var( $u_{k2}$ )	0.007931	0.0011	7.47	<0.001
cov( $u_{k1}, u_{k2}$ )	0.001166	0.0002	6.17	<0.001
$\sigma^2$ (error variance)	1.837	0.120	15.37	<0.001

$h=18.6$  m were estimated. When only  $d$  and  $h$  are known, the mean response ( $d_{ipredfixed}$ ) for any height on the stem can be calculated by the use of Eq. 1 and the fixed-effects parameter estimates presented in Table 3. If there is an additional measurement of an upper stem diameter (e.g.,  $d_i=26.0$  cm at  $h_i=6$  m), the random-effects parameters can be estimated by calibration. With this additional measurement and the estimates of the components of the variance–covariance matrix of the random-effects parameters (Table 3), the model is calibrated as follows:

$$q(\text{relative height at 6 m}) = h_i/h = 0.3226$$

$$p = 1.3/h = 0.06989$$

$$w = 1 - q^{1/3} = 0.3142$$

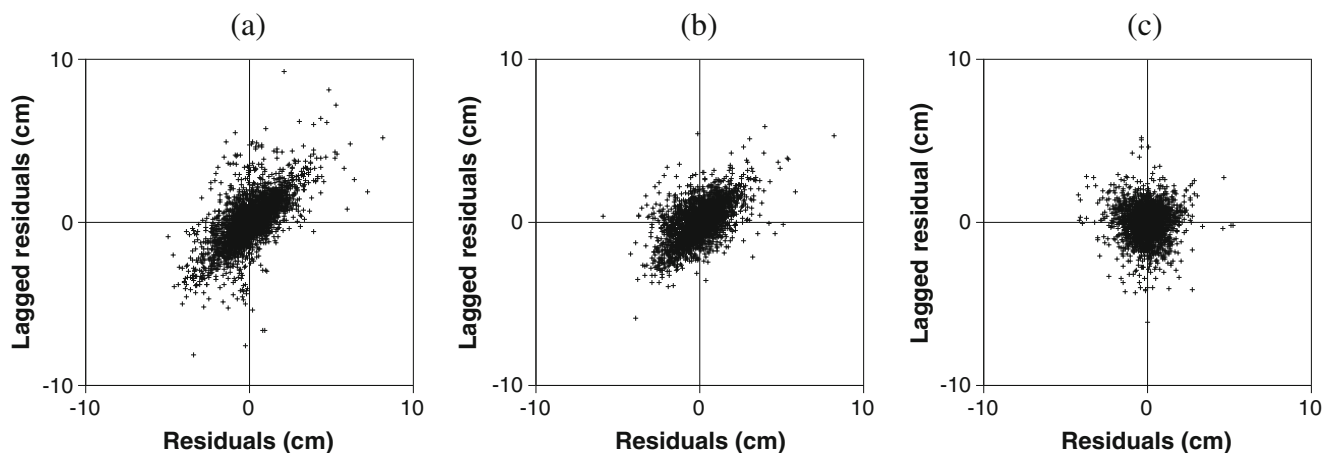
$$x = w/(1 - p^{1/3}) = 0.5342$$

$d_{ipredfixed6m}=24.31$  cm (this is calculated with only the fixed-effects parameters, i.e., it is the mean response)

$$\hat{e}_{ki} = d_i - d_{ipredfixed} = 26.0 - 24.31 = 1.69 \text{ cm}$$

$$Z_{u_{k1}} = a_0 d^{a_1} h^{a_2} x^{b_0 + b_1 q^4 + b_2 (1/\exp(d/h)) + b_3 x^{0.1} + b_4 (1/d) + b_5 h^w + b_6 x} \log(d)$$

(partial derivative with respect to random-effect parameter  $u_{k1}$ )



**Fig. 2** Diameter over bark residuals plotted against lagged residuals for Eq. 8 fitted **a** only with fixed-effects parameters, **b** with fixed- and random-effects parameters, and **c** with fixed- and random-effects parameters plus an autoregressive error structure AR (1)

$$Z_{u_{k2}} = a_0 d^{a_1} h^{a_2} x^{b_0 + b_1 q^4 + b_2 (1/\exp(d/h)) + b_3 x^{0.1} + b_4 (1/d) + b_5 h^w + b_6 x} \log(x)$$

(partial derivative with respect to random-effect parameter  $u_{k2}$ )

$$\hat{\mathbf{Z}}_k = [Z_{u_{k1}} \quad Z_{u_{k2}}] = [85.24 \quad -15.24]$$

$$\hat{\mathbf{R}}_k = \sigma^2 \times \mathbf{\Gamma}_k(\rho) = 1.837 \times [1] = 1.837$$

$$\hat{\mathbf{D}} = \begin{bmatrix} 0.0003373 & 0.001166 \\ 0.001166 & 0.007931 \end{bmatrix}$$

The random-effects parameters are estimated by the use of Eq. 4:  $\hat{\mathbf{u}}_k = (0.005966 \quad -0.01169)^T$ . These estimates of the random-effects parameters can be used in the specific model for the tree in the example by using Eq. 8 with the random-effects parameters  $u_{k1}$  and  $u_{k2}$  as follows:

$$d_{ipred} = 0.9652d^{(0.9421+0.005966)} h^{0.08482} x^c,$$

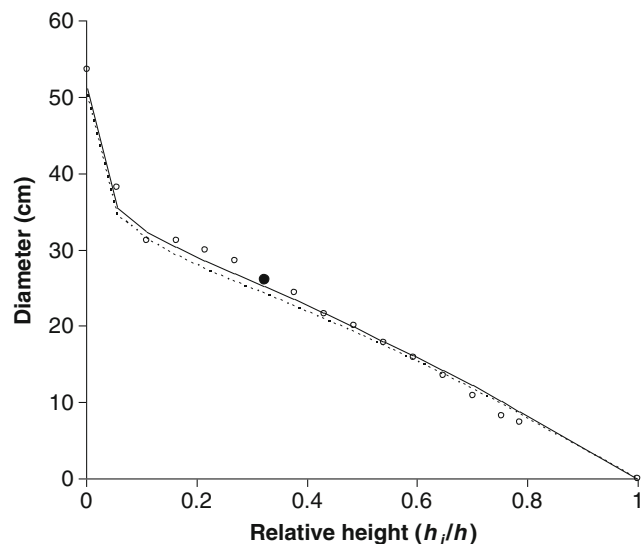
Where

$$c = (4.094 - 0.01169) - 0.3939q^4 - 0.4112 \left( \frac{1}{\exp\left(\frac{d}{h}\right)} \right) - 4.153x^{0.1} + 2.813 \left( \frac{1}{d} \right) + 0.05345h^w + 0.3332x$$

For this example, the mean response and the calibrated response are shown in Fig. 3.

#### 4 Discussion

Single taper models have been found to represent stem shape quite accurately (e.g., Bruce et al. 1968; Kozak et al. 1969;



**Fig. 3** Observed diameters (circles), mean response (dashed line) and calibrated response (solid line) for a tree with diameter at breast height  $d=33.3$  cm and total height  $h=18.6$  m. The additional data used for calibration ( $d_i=26.0$  cm at  $h_i=6$  m) is represented by a black dot

Goulding and Murray 1976), although more flexible models were introduced (Max and Burkhardt 1976; Kozak 1988) in an attempt to provide a better description of the stem profile, especially in the high-volume butt region (Cao et al. 1980). Mixed models have also been used successfully in the development of taper equations for several species throughout the world (e.g., Garber and Maguire 2003; Leites and Robinson 2004; Trincado and Burkhardt 2006; Yang et al. 2009), showing that estimates of the mean response (fixed-effects model) can be significantly improved by including random-effects parameters. This was confirmed in the present study by using the modification of Kozak's (2004) model proposed by Yang et al. (2009).

The mixed model in which the fixed-effects parameters  $a_1$  and  $b_0$  were expanded with the corresponding random-effects parameters  $u_{k1}$  and  $u_{k2}$  was finally selected (Eq. 8) because it had the lowest MSE up to 5 m of calibration height, and it worked well for all calibration heights between 2 and 8 m. When the diameter used in calibration was measured higher in the stem, the MSE of the tested calibrated models was lower (Fig. 1), but adequate calibration was achieved with a diameter at a height between 4–6 m. Equation 8 was developed to be calibrated, but the mean response provided similar accuracy (MSE) as Eq. 1 fitted by ONLS and correction for autocorrelation (Fig. 1). The combination of expanded parameters  $b_0$  and  $b_6$ , which was the best as regarding fitting statistics, showed the lowest MSE for calibration heights above 6.5 m (Fig. 1). As already explained, this model is not recommended from a practical point of view because if calibrations heights lower than 6 m are used, poorer results would be obtained than with Eq. 8. Moreover, the MSE for the mean response of this model was  $2.46 \text{ cm}^2$ , greater than in the mixed model selected ( $\text{MSE}=2.44 \text{ cm}^2$ ).

Although some authors have tried to improve predictions of taper equations by including variables related to the dimensions of the crown, tree age, and other individual tree or stand variables (e.g., Burkhardt and Walton 1985; Muhairwe et al. 1994; Tasissa and Burkhardt 1998), it was generally concluded that as only slight improvements were achieved, the additional effort required to measure these variables was unjustifiable. For example, the  $d/h$  ratio (included in the model of Kozak 2004) is closely related to crown size and type and to stand variables (Burkhardt and Walton 1985; Hann et al. 1987) and the inclusion of random-effects parameters in the model is an effective procedure for capturing the variation that can be explained by these variables.

Although mixed models can be used to model heterocedasticity and autocorrelation (Pinheiro and Bates 2000; Yang et al. 2009), there is a trade-off between the number of random effects incorporated in the model and the complexity of the error covariance structure within each tree (Jones 1990; Pinheiro and Bates 2000). Some studies have shown that modeling the random effects with a

variance–covariance matrix structure can explain the correlation between observations (Vonesh and Chinchilli 1997; VanderSchaaf and Burkhart 2007). However, other studies have shown that although the autocorrelation can be reduced by the use of random-effects parameters, it cannot be completely eliminated (Garber and Maguire 2003; Trincado and Burkhart 2006; Yang et al. 2009). Therefore, in the latter case, the autocorrelation within each individual must be modeled directly. In the present study, the inclusion of random effects in the model partly corrected the autocorrelation (Fig. 2b), while the inclusion of a first order autoregressive structure AR (1) in the mixed model completely removed the autocorrelation (Fig. 2c). Accounting for autocorrelation does not improve the predictive ability of the model, but it prevents underestimation of the covariance matrix of the parameters, thereby making it possible to carry out the usual statistical tests (West et al. 1984), i.e., it improves interpretation of the statistical properties. The model estimations were not significantly different from those obtained with models fitted without considering such correction. The autocorrelation parameter is not used in practical applications unless the model is calibrated with several diameter measurements at different heights on the same tree.

Taper equations can provide volume estimates to various top diameter limits when integrated. If the main aim is to estimate tree volumes, it is preferable to develop separate volume and volume ratio equations (Burkhart 1977; Cao and Burkhart 1980) or to use taper equations with  $d$  squared as the independent variable (Gregoire et al. 2000). However, for estimating merchantable volumes, the first step is to establish the heights at which certain top diameter limits are reached, which allows classification of the stem by merchantable sizes or limits, according to different industrial destinations (peeling, sawing, pulp wood, etc.). The best solution is for the taper equation to have a close-form integral and a generalized inverse,  $h_i=f(d_i, d, h)$ . However, the modified model of Kozak (2004) fitted in this study cannot be directly integrated and does not have a generalized inverse. Numerical integration methods and iterative procedures for estimating the height to any specific diameter are necessary for this purpose.

Trees that do not have deformed stems are used in the development of taper equations. However, the models developed may provide less accurate estimates if applied to trees with different stem types. Therefore, the use of data from deformed trees (e.g., forked or lacking apical dominance) is recommended for developing the most appropriate taper equations for different stem types (e.g., Adu-Bredu et al. 2008), in a similar way as done in the Spanish National Forest Inventory (ICONA 1993; DGCONA 2002).

## 5 Conclusions

A mixed-effects taper equation, based on the model of Kozak (2004), was developed for *Betula pubescens* in Galicia. Both the mean and the specific responses of each individual were obtained. The mean response takes into account only fixed-effects parameters, while the specific response for each individual adds random-effects parameters. In a new individual, the random-effects parameters can be estimated by the use of an approximate Bayesian estimator to provide a calibrated response, which needs a complementary diameter measurement. Results showed that calibration improved model estimations.

Almost two-thirds of the mixed models generated by expansion with one or two random-effects parameters were rejected because they included nonsignificant parameters. The autocorrelation was partly corrected by the use of the mixed-effects approach, but an autoregressive error structure AR (1) was required for the total elimination of autocorrelation.

Because the calibrated mixed model is intended for practical use, calibration should be taken into account in mixed model selection. Selection of mixed-effects parameters based only on fitting statistics may not provide the best calibration statistics for all upper stem diameter measurements. The methodology proposed in this study should be evaluated with other species and different models for comparison of the results.

**Funding** The present study was financially supported by the Spanish Ministry of Education and Science through the research project *Modelos de evolución de bosques de frondosas caducifolias del noroeste peninsular* (AGL2007-66739-C02-01/FOR), co-funded by the European Union through the European Regional Development Fund.

## References

- Adu-Bredu S, Bi AFT, Bouillet JP, Mé MK, Kyei SY, Saint-André L (2008) An explicit stem profile model for forked and unforked teak (*Tectona grandis*) trees in West Africa. For Ecol Manage 255:2189–2203
- Avery TE, Burkhart HE (2002) Forest measurements, 5th edn. McGraw-Hill, New York
- Bi HQ (2000) Trigonometric variable-form taper equations for Australian eucalyptus. For Sci 46:397–409
- Bruce D, Curtis RO, Vancoevering C (1968) Development of a system of taper and volume tables for red alder. For Sci 14:339–350
- Burkhart HE (1977) Cubic-foot volume of loblolly pine to any merchantable top limit. South J Appl For 1:7–9
- Burkhart HE, Walton SB (1985) Incorporating crown ratio into taper equations for loblolly pine trees. For Sci 31:478–484
- Cao QV, Burkhart HE (1980) Cubic foot volume of loblolly pine to any height limit. South J Appl For 4:166–168



- Cao QV, Burkhardt HE, Max TA (1980) Evaluation of two methods for cubic–volume prediction of loblolly pine to any merchantable limit. *For Sci* 26:71–80
- Castroviejo S, Laínz M, López González G, Monteserrat P, Muñoz Garmendia F, Paiva J, Villar L (1990) Flora ibérica: Plantas vasculares de la Península Ibérica e Islas Baleares. Volumen II: Platanaceae-Plumbaginaceae (partim.). RJB (CSIC), Madrid
- Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL (1983) Timber management: a quantitative approach. Krieger Publishing Company, New York
- Cochran WG (1963) Sampling techniques, 2nd edn. Wiley, New York
- Corral-Rivas JJ, Diéguez-Aranda U, Corral Rivas S, Castedo Dorado F (2007) A merchantable volume system for major pine species in El Salto, Durango (Mexico). *For Ecol Manage* 238:118–129
- Davidian M, Giltinan DM (1993) Some general estimation methods for nonlinear mixed-effects models. *J Biopharm Stat* 3:23–55
- Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman & Hall, New York
- DGCONA (2002) Tercer Inventario Forestal Nacional, 1997–2006: Galicia. Dirección General de Conservación de la Naturaleza. Ministerio de Medio Ambiente, Madrid
- Diéguez-Aranda U, Castedo-Dorado F, Álvarez-González JG, Rojo A (2006) Compatible taper function for Scots pine plantations in northwestern Spain. *Can J For Res* 36:1190–1205
- Fang Z, Bailey RL (2001) Nonlinear mixed-effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For Sci* 47:287–300
- Fang Z, Borders BE, Bailey RL (2000) Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *For Sci* 46:1–12
- Fonweban J, Gardiner B, Macdonald E, Auty D (2011) Taper functions for Scots pine (*Pinus sylvestris* L.) and Sitka spruce (*Picea sitchensis* (Bong.) Carr.) in Northern Britain. *Forestry* 84:49–60
- Garber SM, Maguire DA (2003) Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. *For Ecol Manage* 179:507–522
- Goulding CJ, Murray JC (1976) Polynomial taper equations that are compatible with tree volume equations. *N Z J For Sci* 5:313–322
- Gregoire TG, Schabenberger O, Kong F (2000) Prediction from an integrated regression equation: a forestry application. *Biometrics* 56:414–419
- Hann DW, Walters DK, Scrivani JA (1987) Incorporating crown ratio into prediction equations for Douglas-fir stem volume. *Can J For Res* 17:17–22
- Hartford A, Davidian M (2000) Consequences of misspecifying assumptions in nonlinear mixed effects models. *Comput Stat Data Anal* 34:139–164
- ICONA (1993) Segundo Inventario Forestal Nacional. Ministerio de Agricultura, Pesca y Alimentación, Madrid
- Jones RH (1990) Serial correlation or random subject effects? *Commun Stat Simul Comput* 19:1105–1123
- Kozak A (1988) A variable–exponent taper equation. *Can J For Res* 18:1363–1368
- Kozak A (2004) My last words on taper functions. *For Chron* 80:507–515
- Kozak A, Munro DD, Smith JHG (1969) Taper functions and their application in forest inventory. *For Chron* 45:278–283
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38:963–974
- Leites LP, Robinson AP (2004) Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. *For Sci* 50:204–212
- Lindstrom MJ, Bates DM (1990) Nonlinear mixed-effects models for repeated measures data. *Biometrics* 46:673–687
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS for mixed models, 2nd edn. SAS Institute Inc., Cary, NC
- Max TA, Burkhardt HE (1976) Segmented polynomial regression applied to taper equations. *For Sci* 22:283–289
- Muhairwe CK (1999) Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia. *For Ecol Manage* 113:251–269
- Muhairwe CK, LeMay VM, Kozak A (1994) Effects of adding tree, stand, and site variables to Kozak's variable–exponent taper equation. *Can J For Res* 24:252–259
- Newnham RM (1988) A variable-form taper function. Information Report PI-X-83. Petawawa National Forest Institute. Canadian Forest Service, Petawawa, Ontario, Canada
- Newnham RM (1992) Variable-form taper functions for four Alberta tree species. *Can J For Res* 22:210–223
- Pérez D, Burkhardt HE, Stiff C (1990) A variable-form taper function for *Pinus oocarpa* Schiede. in Central Honduras. *For Sci* 36:186–191
- Pinheiro JC, Bates DM (1995) Approximations to the log-likelihood function in the nonlinear mixed effects model. *J Comput Graph Stat* 4:12–35
- Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York
- SAS support (2011) Sample 25032: %NLINMIX macro to fit nonlinear mixed models. <http://support.sas.com/kb/25/032.html>. Accessed 7 July 2011
- Tasissa G, Burkhardt HE (1998) An application of mixed effects analysis to modeling thinning effects on stem profile of loblolly pine. *For Ecol Manage* 103:87–101
- Trincado G, Burkhardt HE (2006) A generalized approach for modeling and localizing stem profile curves. *For Sci* 52:670–682
- Valentine HT, Gregoire TG (2001) A switching model of bole taper. *Can J For Res* 31:1400–1409
- VanderSchaaf CL, Burkhardt HE (2007) Comparison of methods to estimate Reineke's maximum size–density relationship species boundary line slope. *For Sci* 53:435–442
- Vonesh EF (1996) A note on the use of Laplace's approximation for nonlinear mixed effects models. *Biometrika* 83:447–452
- Vonesh EF, Chinchilli VM (1997) Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker Inc, New York
- West PW, Ratkowsky DA, Davis AW (1984) Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. *For Ecol Manage* 7:207–224
- Wolfinger RD, Lin X (1997) Two Taylor-series approximation methods for nonlinear mixed models. *Comput Stat Data Anal* 25:465–490
- Xunta de G (2001) O monte galego en cifras. Dirección Xeral de Montes e Medio Ambiente Natural. Consellería de Medio Ambiente. Santiago de Compostela (Spain)
- Yang Y, Huang S, Trincado G, Meng SX (2009) Nonlinear mixed-effects modeling of variable–exponent taper equations for lodgepole pine in Alberta, Canada. *Eur J For Res* 128:415–429