

Stand and tree-level variability on stem form and tree volume in *Pinus pinea* L.: A multilevel random components approach

R. Calama* and G. Montero

Centro de Investigación Forestal, CIFOR-INIA. Apdo. 8111. 28080 Madrid. Spain

Abstract

Stem analysis data from 536 sampled trees located in 36 permanent sample plots were used to develop a taper function which allows flexible end products volume estimation for stone pine (*Pinus pinea* L.) in Spain. To alleviate inference problems derived from high correlation among observations a multilevel linear mixed approach, including random coefficients varying at both plot and tree levels, was used. The proposed taper function expresses the section diameter as a function of breast height diameter and relative height of the section, showing a logical behaviour at both breast and total height. Between and within plot stem form variability was explained by including explanatory covariates as basal area or distance independent competition indices.

Mixed models allow calibration of the model for new locations, by predicting random coefficients if additional stem form measurements are available. Several alternatives of calibration, considering section diameters measured at different heights in a variable number of trees per plot, were compared between them, and with the basic marginal model and the marginal model including covariates. The best calibration alternative was to use additional section diameter measurements taken at 0.5 m height above ground. This model results in substantial improvement in stem form and single tree volume predictive ability over previously existing volume functions for the species, allowing size end-use classification for timber products.

Key words: taper function; random coefficients; calibration; multilevel models; *Pinus pinea*

Resumen

Variabilidad a nivel de parcela y de árbol de la forma de fuste y del volumen individual en *Pinus pinea* L.: una aproximación multinivel con componentes aleatorios

En el presente trabajo se han utilizado los datos del análisis de perfil de 536 árboles tipo de la especie *Pinus pinea* L., procedentes de 36 parcelas de experimentación, para desarrollar una función de perfil para la especie que permita la estimación de volúmenes individuales y una clasificación flexible de productos maderables de acuerdo a las dimensiones del fuste. Para reducir los problemas de inferencia estadística asociados con la alta correlación identificada entre las observaciones se ha propuesto la formulación de la función de perfil como un modelo mixto lineal multinivel, incluyendo parámetros aleatorios al nivel de árbol y parcela. La función de perfil desarrollada define el diámetro de la sección en función del diámetro normal y la altura relativa de la sección, y presenta comportamiento lógico en la predicción de diámetro a las alturas normal y total. La variabilidad sistemática en la forma del árbol identificada entre parcelas y entre árboles dentro de una misma parcela queda explicada mediante la inclusión de covariables tales como el área basimétrica o índices de competencia.

Los modelos mixtos permiten calibrar la función de perfil para nuevas localizaciones, al predecir los parámetros aleatorios a partir de un número reducido de mediciones adicionales de la forma de fuste. De entre las distintas alternativas de predicción evaluadas, se considera que la mejor es la calibración utilizando valores del diámetro de sección medido a 0.5 m en nuevos árboles. Este modelo supone una mejora en la predicción de la forma de tronco y del volumen individual respecto de las ecuaciones de cubicación desarrolladas con anterioridad para la especie en España.

Palabras clave: función de perfil; parámetros aleatorios; calibración; modelos multinivel; *Pinus pinea*.

* Corresponding author: rcalama@inia.es

Received: 11-07-05; Accepted: 28-10-05.

Introduction

Volume tables and equations, which estimate the volume for a single tree as a function of diameter at breast height and total height, have been traditionally used to calculate total stand volume. More recently, it has been deemed necessary to make accurate estimations of the partial volume of different portions of the log. The reasons for this are:

- To estimate the volume of available timber of different sizes and qualities in order to determine the possible end-uses for a single log i.e. saw, pulp, plywood (Amidon, 1984; Reed and Green, 1984; Prieto and Tolosana, 1991)

- To determine the accumulated biomass and the period for fixed CO₂ to be returned to the atmosphere for each fraction of the tree, as the size of the portion will influence its end use (Montero *et al.*, 2002).

- To determine the effect of stand variables, tree characteristics and silvicultural treatments on volume for the different product classes (Prieto and Tolosana, 1991; Muhairwe *et al.*, 1994; Tassisa and Burkhart, 1998)

Taper or stem profile functions provide an interesting approach to volume estimation for the different merchantable portions of a tree (Kozak *et al.*, 1969). A taper function relates the diameter at any point on the stem to the height at this point. By integrating the taper function between two given heights it is possible to estimate the total volume contained in that portion of stem, therefore allowing a flexible size classification for the products.

The data used to fit taper functions usually come from sample trees, for which diameter over bark is measured at different heights along the stem. Measurements from the same sample tree tend to be highly correlated (Bruce *et al.*, 1968; Amidon, 1984; Gregoire and Schabenberger, 1996). If different trees are sampled in the same stand, the similarity of the measurements between those trees would be above the average. This lack of independence of the data, on a two-level scale (observation within trees and trees within plots) prevents the use of statistical methods based on ordinary least squares (OLS) techniques (West *et al.*, 1984; Fox *et al.*, 2001), since these techniques would lead to bias in the estimation of the confidence intervals for the parameters.

Mixed models, including both fixed parameters common for the whole population and random parameters specific for each sampling level and unit, approach the problem of correlation among observations (Laird and Ware, 1982; Searle *et al.*, 1992; Vonesh and Chinchilli, 1997). A subjacent idea on mixed models is that the relation between two variables, in taper function diameter and height at a point along the stem, can be explained using a common functional structure with specific parameters for each location or occasion. This idea is supported by the general opinion that stem form varies from stand to stand (Laasasenaho, 1982; Lappi, 1991; Gregoire and Schabenberger, 1996) and that within the same stand, stem form varies from tree to tree (Kilikki, 1983; Tassisa and Burkhart, 1998; Eerikäinen, 2001). The systematic pattern of variability in stem form between trees located in different plots can be explained by including stand variables in the taper function (Thomson and Baarclay, 1984; Lappi, 1990; Muhairwe *et al.*, 1994). Form variability within the same plot can be explained by adding single tree variables or variables indicating social status of the tree (Kilikki, 1983; Kozak, 1988; Newnham, 1992; Valentine and Gregoire, 2001).

The Stone pine (*Pinus pinea* L.) is a typical Mediterranean species. In Spain, in both pure and mixed stands, it occupies about 450.000 ha, which is more than the 50% of the total surface of this species in the world. Stone pine stands play an important role in landscaping and are often found in social and recreational areas. They are also soil protectors, since they occupy sandy areas both inland and in coastal regions. The two most important merchantable products from these stands are pinyons (edible pine nuts) and timber. The volume of stone pine timber harvested in Spain during 1998 was 203,698 m³ (MAPA, 2001). This amount is only a small part of the total volume of timber harvested annually in Spain [15,096,500 m³/year between 1989 and 1998 (MAPA, 2001)]. However, its importance is evident because it is the only timber producing species in many areas. The end-uses for stone pine timber include: saw wood for construction, packing, pallets, pulpwood for chipboard and paper manufacture.

Several volume equations are available for the stone pine in Spain which have national (Pita, 1967; Martínez-Millán *et al.*, 1993) or regional (ICONA, 1990) validity. These equations have been developed

Table 1. Fitting data set summary

Region	Number of plots	Number of trees	dbh (mm)		H (dm)		V (dm ³)	
			Mean	min-max	Mean	min-max	Mean	min-max
West Andalusia	18	270	216	72-452	122	36-248	232.5	8-1978
Central Range	4	60	210	96-294	81	34-123	140.5	21-378
Northern Plateau	14	206	268	90-421	105	42-178	270.2	13-1046

Where dbh is breast height diameter, H is total height and V is tree volume.

to estimate single tree volume from breast height diameter, total height and a form factor or a complementary upper diameter measurement. These equations do not take into account the influence of stand, tree or silvicultural effects, and they do not enable product classification.

The aim of this work is to develop a flexible taper function with logical behaviour for tree volume estimation which allows flexible end-product classification for stone pine in Spain. To avoid inference difficulties derived from high correlation among observations taken on sample trees, the proposed taper function is formulated as a multilevel mixed model, including random coefficients at both tree and stand level. The ability of different stand or tree level covariates to explain variability in stem form was also evaluated, and compared with local calibration of the models by using additional measurements in new locations.

Material

During the sixties the CIFOR-INIA installed a network of permanent plots to study growth and yield for the most important forest species in Spain. From each plot, fifteen sample trees were randomly selected, and breast height diameter, total height and merchantable height to a top diameter of 7 cm for each tree were recorded. Measurements for diameter over bark were taken at stump height, at 0.5 meters, and then at every meter up to a height of 10 meters. From 10 meters up to the merchantable height, measurements were taken at every two meters. Trees were not felled, as measurements were taken by workers who climbed the trees.

For *Pinus pinea*, 36 permanent plots were installed in three different regions: West Andalusia, Central Range and Northern Plateau. Taper measurements were recorded during the second inventory (year 1971). The total number of trees available for fitting the data set, as well as their regional distribution and main characteristics are shown in table 1.

Bias in the data set was detected, stemming from the fact that for two trees with the same diameter and height, the 7 cm section diameter can be measured at different relative heights (depending on the tapering of the stem), indicating that measurements taken on the upper part of the stem (higher relative height sections) are all from the less tapered trees. To avoid this, trees were classified into classes per 2 cm breast height diameter and 1 meter total height. In each class, all the observations coming from a section whose height is equal to or higher than the smallest merchantable height for the class were rejected (figure 1). The total number of available observations after data depuration was 4365.

Methods

Taper function

Since Kozak *et al.* presented their seminal work (1969), a wide range of taper function equations have been developed to predict stem form. These include polynomial, segmented polynomial, exponential, variable exponent, potential or trigonometric equations (see Castedo and Álvarez, 2000 or Novo *et al.*, 2003 for complete recent revisions on taper functions). Despite all these advances, today it is difficult to identify a single taper function or family of taper functions which

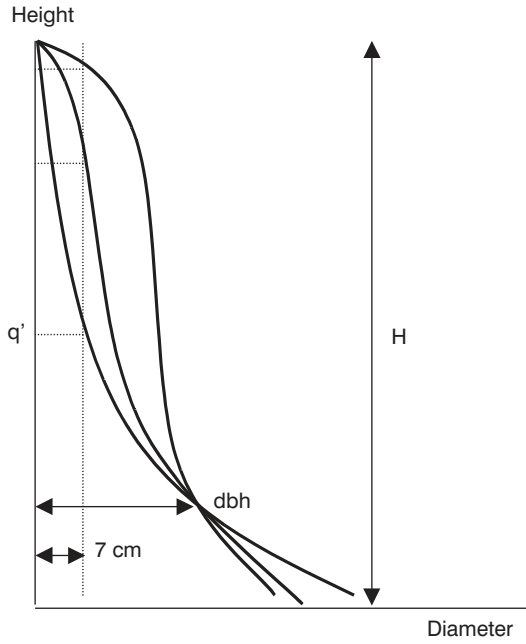


Figure 1. Bias derived from the difference in relative height for the 7 cm section diameter for three trees with the same breast height diameter (*dbh*) and total height (*H*). Observations taken above *q'* are rejected.

present a clear advantage over the rest. Therefore, new taper models are continuously being developed. In the present work, we propose to modify Amidon's classical polynomial taper function (1984). The expression for original model is shown in Eq. 1:

$$d_{ub} = a_1 dbh \left(\frac{H-h}{H-13} \right) + a_2 \left(\frac{(H^2 - h^2)(h-13)}{H^2} \right) \quad (1)$$

Where d_{ub} is the section diameter under bark at height h ; dbh is breast height (13 dm) diameter, and H is total height for the tree. Henceforth, heights are given in dm and diameters in mm. This function presents several advantages, such as its simple formulation to directly predict section diameter, its linear character without transformation, its logical behaviour at total height H (where predicted section diameter d is zero); as well as the possibility for direct integration between two given heights in order to estimate bole volume without bark.

The first restriction found when using this expression arises because we are working with section diameters over bark. This means that if we want a

function with a logical behaviour, we have to include an additional constraint, since predicted section diameter d over bark at breast height 13 dm should equal breast height diameter dbh . Most of the available taper functions do not match this condition. Despite this fact, in an OLS regression context, it is always possible to include some restriction in parameter estimates (generally to identify a parameter as a linear combination of the rest) leading to logical behaviour, or simply to consider that unbiased prediction errors for breast height diameter only slightly affect taper and volume estimations.

In a random coefficient approach things are not so simple, since parameters are not fixed, but rather, mixed (composed of a fixed and a random part), so it is not possible to find a single combination of parameters giving null (or at least low) errors for breast height section. As a result, errors in predicting diameter at breast height will be biased and highly correlated between observations coming from the same plot (since they will share common random plot-level parameters). Trivial correlations would also be identified between random plot and tree parameters. To alleviate this, as a first step we propose a modification of Eq. 1:

$$d_{ijk} = \underbrace{dbh \left(\frac{H_{ij} - h_{ijk}}{H_{ij} - 13} \right)}_{A_{ijk}} + a_1 \underbrace{\left(\frac{(H_{ij}^2 - h^2)(h_{ijk} - 13)}{H_{ij}^2} \right)}_{B_{ijk}} \quad (2)$$

Where d is the section diameter over bark at height h , the rest as in Eq. 1; subindexes ijk indicate k^{th} section from j^{th} tree growing in the i^{th} plot. By removing parameter a_1 from Eq. 1, the taper function now shows a logical behaviour at both $h = 13$ and $h = H$. Equation 2 has a single parameter a_1 , affecting the continuous term B_{ijk} , whose value is zero at $h = 13$ dm and $h = H$, and larger than zero for the rest of the points within the interval. B_{ijk} has a maximum at section height h_{max} , depending on total height H . Consequently, between 13 and H , for any $h < h_{max}$ it is possible to find another $h' > h_{max}$ where the value for B_{ijk} would be the same. This means that the stem equation is not flexible, since predicted section diameters at h and h' always deviate by the same quantity from the value given by term A_{ijk} , in other words, both predictions are not independent. Together with this, term B_{ijk} for h is always negative below 13 and positive for h above 13. This means

that, irrespective of the sign of a_1 , when using Eq. 2, deviations from A_{ijk} are always of the same sign if $h < 13$, and of the contrary sign if $h > 13$, which again leads to high correlation between estimates. To deal with all these limitations, we changed exponents '2' in term B_{ijk} and an additional term C_{ijk} was included in the model.:

$$d_{ijk} = \underbrace{dbh_{ij} \left(\frac{H_{ij} - h}{H_{ij} - 13} \right)}_{A_{ijk}} + a_1 \underbrace{\left(\frac{(H_{ij}^\alpha - h_{ijk}^\alpha)(h_{ijk} - 13)}{H_{ij}^\alpha} \right)}_{B_{ijk}} + a_2 \underbrace{\left(\frac{(H_{ij} - h_{ijk})^\beta (h_{ijk} - 13)}{H_{ij}^\beta} \right)}_{C_{ijk}} \quad (3)$$

In order to maintain the linear character of the model, terms α and β in the equation were fixed heuristically, by comparing different sets of values α

and β in terms of log-likelihood. The best result was obtained considering $\alpha = 1.5$ and $\beta = 4$. The proposed taper function shows similar functional structure and characteristics to the Farrar (1987) or Bennet and Swindel models (1972).

Multilevel linear mixed modelling

To avoid the problems derived from high correlation among observations obtained in the same tree and plot, a multilevel linear mixed model approach (Hox, 1995) was proposed. Equation (3) was formulated as a multilevel random coefficient model, dividing each parameter (a_1, a_2) into a fixed part (β_1, β_2) common for the population, and two random components, one specific to each sampled plot i (u_{1i}, u_{2i} ; with $i = 1$ to 36) and the other specific to each sampled tree j within plot i (v_{1ij}, v_{2ij} ; with $j = 1$ to 15). The basic expression for the multilevel mixed model is given by:

$$d_{ijk} = \underbrace{dbh_{ij} \left(\frac{H_{ij} - h_{ijk}}{H_{ij} - 13} \right)}_{A_{ijk}} + [\beta_1 + u_{1i} + v_{1ij}] \underbrace{\left(\frac{(H_{ij}^{1.5} - h_{ijk}^{1.5})(h_{ijk} - 13)}{H_{ij}^{1.5}} \right)}_{B_{ijk}} + [\beta_2 + u_{2i} + v_{2ij}] \underbrace{\left(\frac{(H_{ij} - h_{ijk})^4 (h_{ijk} - 13)}{H_{ij}^4} \right)}_{C_{ijk}} + e_{ijk} \quad (4)$$

In considering:

$$x_{ijk} = \left[dbh_{ij} \left(\frac{H_{ij} - h_{ijk}}{H_{ij} - 13} \right), \left(\frac{(H_{ij}^{1.5} - h_{ijk}^{1.5})(h_{ijk} - 13)}{H_{ij}^{1.5}} \right), \left(\frac{(H_{ij} - h_{ijk})^4 (h_{ijk} - 13)}{H_{ij}^4} \right) \right]$$

$$z_{ijk} = \left[\left(\frac{(H_{ij}^{1.5} - h_{ijk}^{1.5})(h_{ijk} - 13)}{H_{ij}^{1.5}} \right), \left(\frac{(H_{ij} - h_{ijk})^4 (h_{ijk} - 13)}{H_{ij}^4} \right) \right]$$

$$\beta = [1, \beta_1, \beta_2]^T \quad u_i = \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \quad v_{ij} = \begin{bmatrix} v_{1ij} \\ v_{2ij} \end{bmatrix}$$

It is possible to write the multilevel mixed model as:

$$y_{ijk} = x_{ijk}\beta + z_{ijk} u_i + z_{ijk} v_{ij} + e_{ijk} \quad (5)$$

Where y_{ijk} is the scalar diameter d for the k^{th} section in the j^{th} tree from the i^{th} plot; x_{ijk} is a vector of predictor covariates associated with fixed parameters; β is a vector for fixed parameters; z_{ijk} is a vector of covariates associated with random effects at plot and tree within-plot level; u_i is vector of random parameters operating at plot level, specific for plot i ; v_{ij} is the vector of random parameters operating at tree level, specific for the j^{th} tree within the i^{th} plot; and e_{ijk} is a random error term. The basic assumptions for the multilevel mixed model theory using maximum likelihood estimation methods include the multivariate normal distribution for random effects at each level and for the residual error term:

$$\begin{aligned} u_i &\sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D_i = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u12} & \sigma_{u2}^2 \end{bmatrix} \right] \\ v_{ij} &\sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D_{ij} = \begin{bmatrix} \sigma_{v1}^2 & \sigma_{v12} \\ \sigma_{v12} & \sigma_{v2}^2 \end{bmatrix} \right] \\ e_{ijk} &\sim N(0, \sigma_e^2) \end{aligned}$$

Where D_i and D_{ij} are variance-covariance matrices for random plot and tree components, defining random variability at plot and tree level, respectively. In this work, we consider that D_i and D_{ij} are constant for every plot and tree within the analysed plot, indicating that patterns of correlation between trees within the same plot and between observations within the same tree are constant. Random components acting at different units within the same level (e.g. random parameters specific for different plots) or acting at different levels (e.g. random plot and tree parameters) were considered independent. Model 5 can be rearranged in matricial form to obtain the classical expression for a mixed model (Henderson *et al.*, 1959; Searle, 1971):

$$y = X\beta + Zb + e \quad (6)$$

Where y is the vector including all the section diameters d_{ijk} in the sample (in our case, a 4365×1 vector); X is a matrix for fixed effects, whose rows are vectors x_{ijk} ; Z is a design matrix for random parameters:

$$Z = \begin{pmatrix} Z_1 & & & & \\ & Z_2 & & & \\ & & \cdots & & \\ & & & Z_i & \\ & & & & \cdots \\ & & & & & Z_{36} \end{pmatrix} \text{ where } Z_i = \begin{pmatrix} Z_{i,1} & Z_{i,1} & & & \\ Z_{i,2} & & Z_{i,2} & & \\ \cdots & & & \cdots & \\ Z_{i,j} & & & Z_{i,j} & \\ \cdots & & & & \cdots \\ Z_{i,15} & & & & Z_{i,15} \end{pmatrix} \text{ and } Z_{i,j} = \begin{pmatrix} z_{ij1} \\ z_{ij2} \\ \cdots \\ z_{ijn_j} \end{pmatrix}$$

b is a vector for random parameters, including components at tree and plot level

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_i \\ \cdots \\ b_{36} \end{bmatrix} \text{ with } b_i = \begin{bmatrix} u_i \\ v_{i1} \\ \cdots \\ v_{ij} \\ \cdots \\ v_{i,15} \end{bmatrix}$$

and e is a vector for residual error terms.

Vector b is distributed following a multivariate normal with mean zero and variance-covariance matrix D , where

$$D = \text{Var}(b) = \begin{bmatrix} \text{Var}(b_1) & & & & \\ & \text{Var}(b_2) & & & \\ & & \dots & & \\ & & & \text{Var}(b_i) & \\ & & & & \dots \end{bmatrix}; \text{Var}(b_i) = \begin{bmatrix} D_i & 0 & 0 & \dots & 0 \\ 0 & D_{i1} & 0 & \dots & 0 \\ 0 & 0 & D_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & D_{i15} \end{bmatrix}$$

Vector e is distributed following a multivariate normal distribution with mean zero and variance matrix R . Given this, the first and second moments for the distribution of y are:

$$E(y) = X\beta \quad (7)$$

$$\text{Var}(y) = V = ZDZ^T + R$$

The usual aim of an analysis of this type is the estimation of the vector of fixed effects β , the estimation of the components of the variance-covariance matrices D and R and the prediction of the vector of random effects b . From (7) we can make inference over these parameters using maximum likelihood estimates from the General Linear Theory (Searle, 1971; Laird and Ware, 1982):

$$\begin{aligned} \hat{\beta} &= (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} y \\ \hat{b} &= \hat{D} Z^T \hat{V}^{-1} (y - X\hat{\beta}) \end{aligned} \quad (8)$$

Components of D and R (*variance components*) are generally unknown, but inference can be based on estimates over these matrices (Henderson *et al.*, 1959; Harville, 1977; Searle *et al.*, 1992). Inference over these elements is associated with the maximization of the density function for y , under multivariate normality assumption. To include the loss in degrees of freedom derived from the estimation of the parameters for the fixed effects, the function to be maximized is the marginal restricted likelihood function (REML) (Harville, 1977). Estimation of the variance components was carried out using SAS procedure MIXED (Singer, 1998; Sheu and Suzuki, 2001). Output from this software includes estimates over the variance components, an estimate for the fixed effects vector β and prediction for the vector of empirical best linear unbiased predictors (EBLUP's) for the random effects vector b .

Goodness of fit criteria tests

Akaike's Information Criterion (AIC) has been widely used to compare different mixed models:

$$AIC = -2LL + 2q$$

Where $-2LL$ is -2 times the value for the logarithm of the likelihood function and q is the number of variance components and fixed parameters to be estimated. Smaller values of AIC indicate better data fit. Comparisons between models including different sets of nested fixed effects were also carried out using likelihood ratio test over the value of $-2LL$. Modelling efficiency for the fixed part of the model was also used in comparing models including different subsets of fixed parameters.

Within tree variance-covariance structure

The R matrix explains variability between observations at within tree level. It is usually a diagonal matrix with components σ_e^2 . In a first step, we propose to consider this structure to fit equation (4). This simple structure might not be accepted if residuals from the same tree are not independent at tree-level or if residual variance is not constant (heteroscedasticity).

Lack of independence means that residuals from adjacent observations located in the stem show an above average similarity between them. Heteroscedasticity indicates that accuracy depends on the part of the stem where prediction is carried out. In our case, correlation among residuals from the same tree was tested by plotting the residual against residuals from the observation located just below it on the stem. On the other hand, heteroscedasticity was tested by computing the mean value for the residual variance (using residuals from model 4) for different classes of relative height. The patterns of heteroscedasticity or correlation detected would be considered in the model through variance function and special structures for matrix R .

Results

Fitting basic multilevel linear mixed model

The first column in table 2 shows the fitting statistics for equation 4 considering independence between residuals and homocedastic residual variance. High, significant positive correlation ($p < 0.001$) was identified between residuals delayed one, two or even more positions within the stem. Therefore, we evaluated different correlation structures for matrix R (Gregoire, 1987), selecting a power type one, where correlation between two sections k_1 and k_2 from the same tree is defined by $\gamma^{d_{12}}$, where d_{12} indicates the distance (in dm) between sections k_1 and k_2 , and γ is the power parameter ($\gamma < 1$). The second column in table 2 shows fitting statistics for equation 4 after considering this autocorrelative structure. The large decrease on the $-2LL$ statistic indicates that this structure is a significant improvement on the previous model.

Figure 2 shows a clear pattern of increasing residual variance with greater relative height, indicating that predictions are less accurate for the upper part of the

stem. A cubic function for residual variance ρ_{ijk} was then fitted using OLS linear regression:

$$\rho_{ijk} = 2143 \left[\frac{h_{ijk}}{H_{ij}} \right]^3 - 1425 \left[\frac{h_{ijk}}{H_{ij}} \right]^2 + 177 \left[\frac{h_{ijk}}{H_{ij}} \right] + 104$$

After considering the correlation structure and heterocedastic pattern, the within-tree residual variance-covariance matrix is given by:

$$R = \sigma_e^2 G^{1/2} \Gamma G^{1/2} \quad (9)$$

Where σ_e^2 is the variance for the homocedastic and independent residuals, acting as a scaling factor for the error dispersion (Davidian and Giltinan, 1993; Gregoire *et al.*, 1995); G is a diagonal matrix, with dimension equal to the number of observations, whose components are given by the predicted variance function ρ_{ijk} specific for each observation, and Γ is a matrix describing the correlation pattern between observations coming from the same tree, with values 1 in diagonal and $\gamma^{d_{12}}$ for the rest of components.

Table 2. Fixed parameters, variance components estimates and fitting statistics comparison between different alternatives of weighting, stand and tree covariates inclusion. Model 6 shows results for OLS fitting of the model without covariates

	1	2	3	4	5	6
β_1	1.0972	1.1228	1.1924	1.2121	0.3900	1.3339
β_2	-2.8505	-2.3500	-2.4463	-1.2312	-1.1817	-2.6938
Basal Area (λ_2)	—	—	—	-0.0458	-0.0466	—
d/dg (λ_1)	—	—	—	—	0.8034	—
σ_{u1}^2	0.4943	0.3774	0.3707	0.3566	0.3697	—
σ_{u2}^2	2.8505	2.2945	2.2413	1.6042	1.6298	—
σ_{u12}	-1.0724	-0.8316	-0.7912	-0.6884	-0.7018	—
σ_{v1}^2	0.1436	0.0818	0.041	0.0408	0.0269	—
σ_{v2}^2	-0.2582	0.1338*	—	—	—	—
σ_{v12}	0.6238	-0.1232	—	—	—	—
Weighting factor	1	1	ρ_{ijk}	ρ_{ijk}	ρ_{ijk}	—
Power function γ	—	0.8891	0.9094	0.9097	0.9103	—
σ_e^2	161.16	190.98	1.1121	1.1130	1.1164	267.04
$-2LL$	35405	35031	34142	34129	34083	36785
AIC	35418	35047	34154	34141	34095	36787
EF%	95.39	95.44	95.59	95.76	95.83	95.70

Where EF: modelling efficiency for the marginal (fixed effects only) model, rest of parameters as pointed out in the text. All the parameters associated with both fixed and random effects are significant at $\alpha = 0.05$ except the one indicated with *.

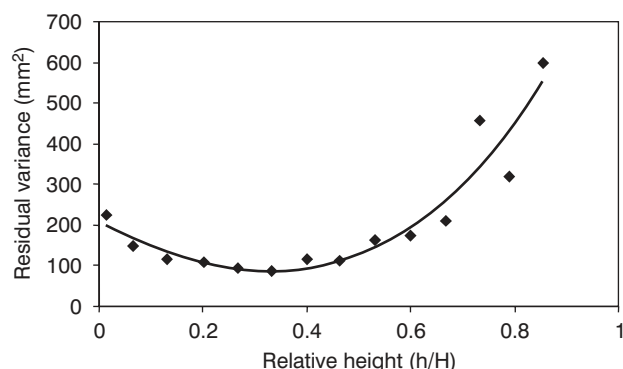


Figure 2. Values for the mean residual variance as a function of relative height (dots), and fitted variance function (solid line).

Equation (4), under variance-covariance assumptions defined in (9), was fitted to the data set using SAS procedure MIXED. Variance components associated with β_2 at tree level ($\sigma_{v_2}^2$ and $\sigma_{v_{12}}^2$) were identified as nonsignificant, and so random parameter v_{2ij} was removed from equation 4. Results for the parameter estimates and goodness of fit statistics are included in the third column of table 4. Comparing the value for the remanent variance components acting at plot and tree level it is possible to affirm that the pattern of variability in stem form identified between trees located in different plots is greater than that identified between trees within the same plot.

Volume estimation

The proposed taper function (4) can be directly integrated between two given section heights to obtain bole volume, since the value for a given squared section diameter d_{ijk}^2 can be expressed as a polynomial form over section height h_{ijk} . Despite this, integration involves high-order terms, so the volume expression becomes quite complex. To avoid this, it is preferable to estimate volume by numerical integration of the function, directly estimating the section diameter and area at constant height intervals along the stem (for example, 5 cm), and computing the volume attained between those height intervals using Smalian's formula (Prodan *et al.*, 1997). Total tree volume is then computed as the sum of those partial volumes, always taking into account that Smalian's formula tends to overestimate volume predictions.

To determine saw timber volume within a given tree, it is necessary to estimate the height at which the tree reaches the minimum diameter for sawing. As h_{ijk} can not be rearranged from equation (4) interpolation techniques are then required to estimate limit heights and size classification of products (as stated, e.g., in Kozak, 1988). A simpler approach on a given tree, is to simulate its division into different logs of commercial length, then estimate the diameter for the upper and lower sections of the log, and assign the end use of the log depending on the upper diameter.

Improving the model

The proposed model can be useful to predict stem form and volume for a given tree, if both breast height diameter and total height of the tree are available. If no additional data are available, predictions will be carried out considering only the fixed part of the model, i.e., considering vectors u_i and v_{ij} as null vectors, or, what is the same, considering random parameters at plot and tree levels to be zero. Under these conditions, we will obtain *marginal prediction*, obtaining the same stem form and volume for every tree having the same breast height diameter and total height.

The inclusion of random components at plot and tree level indicates that a pattern of systematic non-explained variability exists, associated with observations coming from those levels, in other words, that trees with the same size do not necessarily share the same stem form and volume. Due to the stochastic structure of the model, random variability can be explained (at least in part) using two different approaches:

Covariate modelling

Systematic variability in stem form detected between and within plots can be explained in the first place by including both stand and single tree variables characterizing these factors as fixed effects in the model (Muhairwe *et al.*, 1994).

To determine which variables to include in the model, first, the predicted EBLUP's after fitting model 4 were linearly expanded over plot and tree variables (Laird and Ware, 1982). The plot variables tested were: age, dominant height H_0 , basal area BA , mean square diameter dg , number of stems/ha N and site index SI

(obtained from the model by Calama *et al.*, 2003), as well as its logarithmic transformations. Tree level EBLUP's were expanded over variables indicating the status of the tree within the stand: ratio between tree breast height diameter and plot mean square diameter dbh/dg , ratio between tree section and plot basal area g/BA and basal area of the trees larger than subject tree BAL . Those covariates showing high correlation with EBLUP's (figure 3) were then included into the basic model as fixed effects in equation (4), evaluating different alternatives of covariate inclusion on the basis of AIC. No significant

correlation was detected between any covariate and the random component u_{1i} .

The fourth and fifth columns in table 2 show the fitting statistics, fixed parameters and variance components for the model showing best performance after including plot and tree covariates in the basic multilevel mixed model. The expanded models indicate a significant decrease in the $-2LL$ statistic and in the value for the variance components acting at plot and tree levels, together with an increment in the efficiency for the fixed part of the model. The taper function finally proposed is shown in eq. 10:

$$d_{ijk} = dbh_{ij} \left(\frac{H_{ij} - h_{ijk}}{H_{ij} - 13} \right) + \left[\beta_1 + \lambda_1 \frac{dbh_{ij}}{dg_i} + u_{1i} + v_{1ij} \right] \left(\frac{(H_{ij}^{1.5} - h_{ijk}^{1.5}) (h_{ijk} - 13)}{H_{ij}^{1.5}} \right) + [\beta_2 + \lambda_2 BA + u_{2i}] \left(\frac{(H_{ij} - h_{ijk})^4 (h_{ijk} - 13)}{H_{ij}^4} \right) + e_{ijk} \quad (10)$$

The complete model (10) can be used to predict stem form and volume in a similar way to that in which the basic model was used, by giving zero value to u_i and v_{ij} , defining the *marginal covariate prediction*. Covariate modelling allows us to obtain different stem form and volume estimates for trees which, having the same breast height diameter and total height, grow in different conditions (Fig. 4).

Calibration

The stochastic character of model 4 allows us to improve the predictions for new unsampled locations by predicting random components specific to those locations. If complementary measurements of stem form are available, together with breast height diameter and total height, EBLUP's for random plot and tree

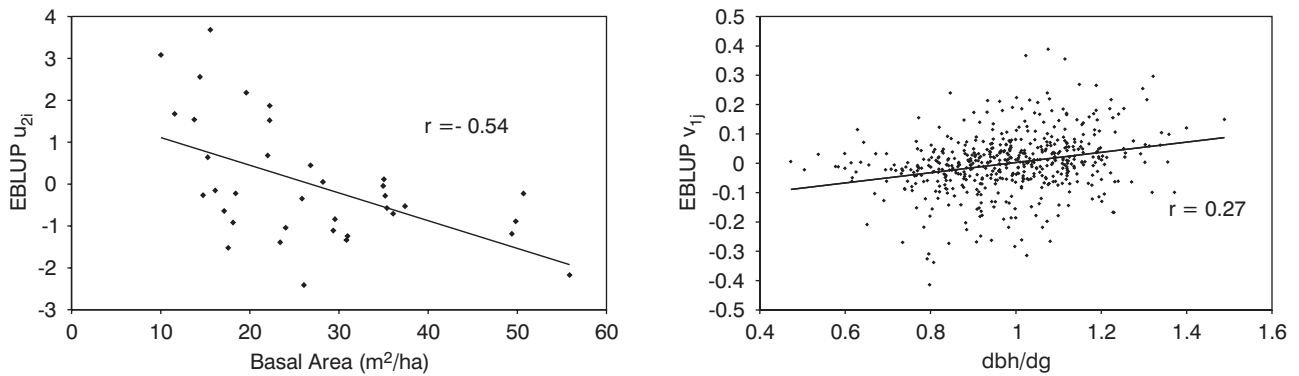


Figure 3. Relation between EBLUP's u_{2i} (left) and v_{1ij} (right) with basal area and dbh/dg .

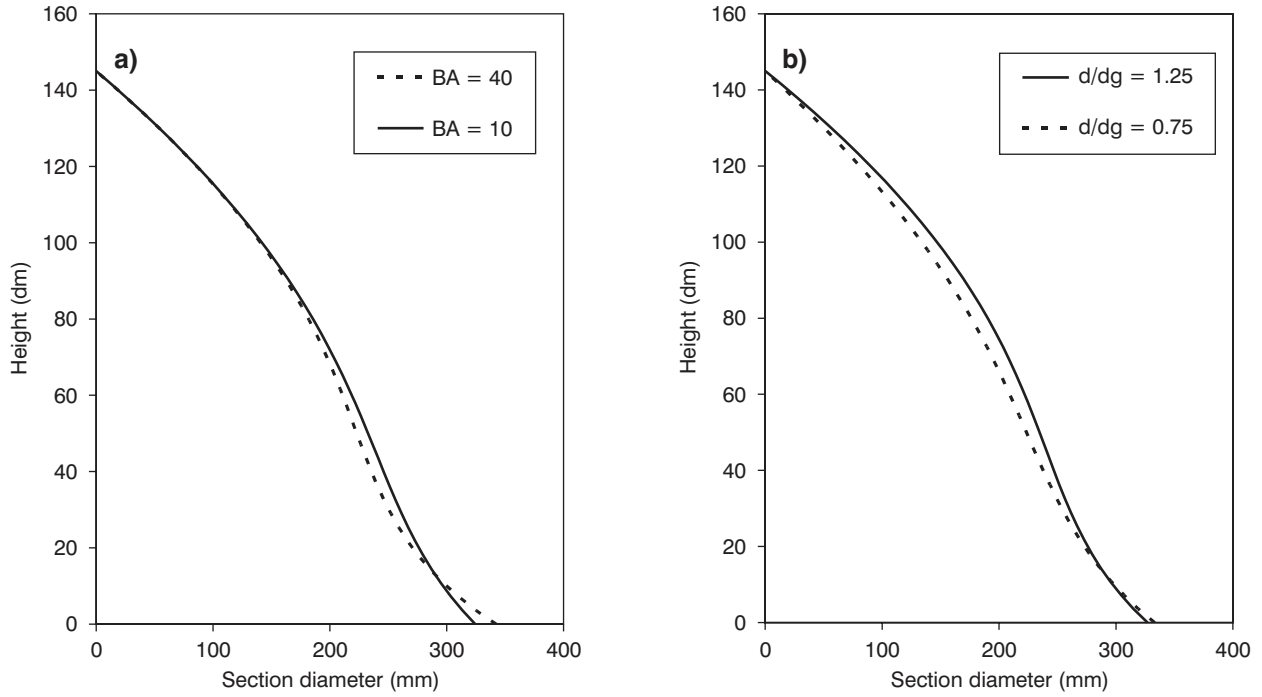


Figure 4. Tree profiles from eq. 10 for $dbh = 290$ mm and $H = 145$ dm a) at different basal area (left) and b) in stands with $BA = 20$ m²/ha, with different social status (right).

parameters can then be predicted. In this case, the model is calibrated for each sampling unit (Bondeson, 1990; Lappi, 1991; Vonesh and Chinchilli, 1997). For a new, additional observation, the vector of random parameters $\hat{b}$ can be predicted using eq. (11):

$$\hat{b} = \hat{D}Z^T(Z\hat{D}Z^T + \hat{R})^{-1}(y - X\hat{\beta}) \quad (11)$$

Where $y - X\hat{\beta}$ shows the differences between additional observed section diameters and the marginal prediction for these observations; X , Z are design matrices associated with the new set of observations; $\hat{D}$ and $\hat{R}$ are predicted variance matrices associated with the new observations. Predictions carried out by using model 4 after including predicted EBLUP's for $\hat{b}$ are defined as *conditional calibrated predictions*.

Case study of calibration

The following is a simple case of calibration by way of explanation:

If we take tree no. 5 from plot 47023, with a dbh equal to 370 mm and total height of 164 dm, by using the fixed part of eq. (4), it is possible to obtain a marginal prediction for stem form and compare it with real known data. Along with this data, if we know that there is a tree (no. 67) within the same plot which has a dbh of 381 mm, total height of 168 dm, and a section diameter measurement at height 45 dm of 341.5 mm, it is possible to predict EBLUP's for random plot parameters $u_{1,47023}$ $u_{2,47023}$ using this latter observation and eq. 11, where, in this case:

$$b = [u_{1,47023}, u_{2,47023}, v_{1,47023,67}]^T; \hat{D} = \begin{bmatrix} 0.3707 & -0.7912 \\ -0.7912 & 2.2413 \\ & & [0.041] \end{bmatrix}$$

$$Z = [27.563 \quad 9.194 \quad 27.563]; \hat{R} = 93.631 \times 1.1121; y - X\hat{\beta} = 341.5 - 312.7 = 28.8$$

After solving eq. 11 we obtain:

$$\hat{u}_{1,47023} = 0.4127; \hat{u}_{2,47023} = -0.1684$$

We include these predicted values in eq (4) in order to estimate stem form and volume for tree no. 5. Random tree component $\hat{v}_{1,47023,67}$ is of no use in calibrating stem form function for tree no. 5, since it has been predicted and defined for tree no. 67. As can be seen in figure 5, calibration reduces both deviation in section diameter and tree volume estimation.

Comparing the predictive ability of the models

Marginal prediction (obtained using the fixed part of eq. 4), marginal covariate prediction (from the fixed part of eq. 10) and conditional calibrated prediction were compared over a subsample of the fitting data set, as an independent validation data set was not available. Only those plots where a diameter measurement was

available at 45 dm for all sample trees were selected (fifteen in all). The evaluation set was made up of 225 trees from the 15 plots. We compared different calibration alternatives at plot level, by predicting random plot components using the following additional measurements

- Calibration after measuring section diameter at 45 dm height in 1, 2, 3, 4 or 5 trees per plot.
- Calibration after measuring section diameter at 5 dm height in 1, 2, 3, 4 or 5 trees per plot.

For each calibration alternative and sample size we carried out 500 random realizations. In each realization, trees measured in each plot to predict random components were randomly selected, and predictions were made for the fifteen trees in the plot. Alternatives were compared using the average value (for the 500 realizations) of the root mean squared error (RMSE) and mean error (Bias) for section diameter and tree volume estimates:

$$RMSE(y) = \sqrt{\frac{\sum_i \sum_j \sum_k (y_{ijk} - \hat{y}_{ijk})^2}{n-1}}$$

$$Bias(y) = \frac{\sum_i \sum_j \sum_k (y_{ijk} - \hat{y}_{ijk})}{n}$$

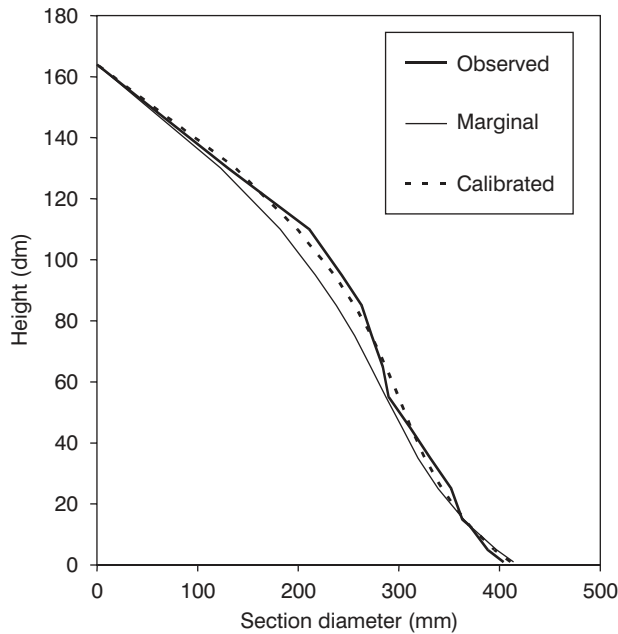


Figure 5. Comparison between observed stem form for tree no. 5, plot 47023 and predicted stem forms using marginal prediction and calibration from an additional observation at 45 dm height (dashed line) taken on a different tree within the plot.

Where y and $\hat{y}$ indicate observed and predicted values for dependent variable (section diameter or tree volume) and n number of observations. The results on the predictive ability of the different alternatives are included in table 3.

As can be seen from table 3, the RMSE for section diameter estimates are not improved after calibration, but bias in section diameter prediction is reduced. In comparing different calibration alternatives, less accurate but less biased predictions are obtained by using additional measurements at 5 dm height. In any case, RMSE for section diameter predictions is between 17 - 18 mm, with a mean error for diameter prediction of between 4.2 mm for the basic fixed effects model and 0.9 mm for the calibrated model using five 5 dm measurements. Bias in diameter predictions is nonsignificant ($p > 0.01$) only after calibrating with five 5 dm additional measurements.

Table 3. Comparison between the stem form predictive ability for the marginal model, the marginal covariate model and different alternatives of plot calibration

	Number of calibrating trees per plot	RMSE (d) mm	Bias (d) mm	RMSE (V) dm ³	Bias (V) dm ³
MARGINAL	0	17.1	3.7	46.6	15.4
MARGINAL + COVARIATE	0	16.8	4.2	44.7	17.7
CALIBRATION ADDITIONAL MEASUREMENT AT 45 dm	1	17.4	3.2	44.3	13.4
	2	17.2	3.0	42.7	12.5
	3	17.1	2.8	41.7	11.8
	4	17.0	2.7	41.0	11.5
	5	16.9	2.7	40.6	11.3
CALIBRATION ADDITIONAL MEASUREMENT AT 5 dm	1	17.1	2.8	44.6	12.5
	2	17.2	2.1	43.6	10.6
	3	17.4	1.6	42.8	9.0**
	4	17.5	1.2	42.1	7.8*
	5	17.6	0.9**	41.9	7.0*

* Nonsignificant with $p > 0.01$; ** Nonsignificant with $p > 0.001$.

With respect to volume predictions, calibration substantially reduces RMSE and bias if compared with marginal models. RMSE for both calibration alternatives are similar, but in using additional observations measured at 5 dm, tree volume predictions are less biased. RMSE for volume prediction ranges from 47 dm³ for the marginal basic multilevel model to 45 dm³ for the marginal covariate model, while calibration alternatives lead to RMSE values under 41 dm³. The mean prediction error ranges from 18 dm³ for the marginal covariate model to values under 10 dm³ for 5 dm calibration alternatives. Marginal predictions tend to underestimate both section diameter and tree volume, while calibration alternatives tend to give more unbiased estimates, especially if additional measurements are taken at 5 dm.

Marginal and calibrated predictions (in the latter case, only using additional measurements from two trees per plot) were also compared by evaluating the trend of the RMSE (average after 500 random realizations in the case of calibrated models) for relative diameter error (12) and relative volume error (13) as a function of relative height (h/H) and total tree volume. To do this, trees from the validation data set were divided into 100 dm³ volume classes, and sections were grouped in 10% relative height classes according to their position on the stem. The use of relative errors

avoids the effect of considering jointly trees and sections of different sizes.

$$\text{Relative diameter error} = \frac{d_{ijk} - \hat{d}_{ijk}}{d_{ijk}} \quad (12)$$

$$\text{Relative volume error} = \frac{V_{ij} - \hat{V}_{ij}}{V_{ij}} \quad (13)$$

RMSE for relative error in section diameter prediction is under 10% up to a relative height of 0.6. From this point to the tip RMSE increases up to 25% (Fig. 6a), indicating less accurate predictions in the upper part of the stem. Slight differences are shown between marginal and calibrated predictions, although calibration using two measurements at 5 dm height per plot leads to larger RMSE values, close to 30%, in relative height over 0.8. With respect to volume prediction (Fig. 6b), RMSE for prediction is under 10% for all volume classes, with calibrated alternatives reaching smaller values only in the largest volume classes (>900 dm³).

Bias in marginal and marginal covariate predictions was analysed by evaluating the mean value of real

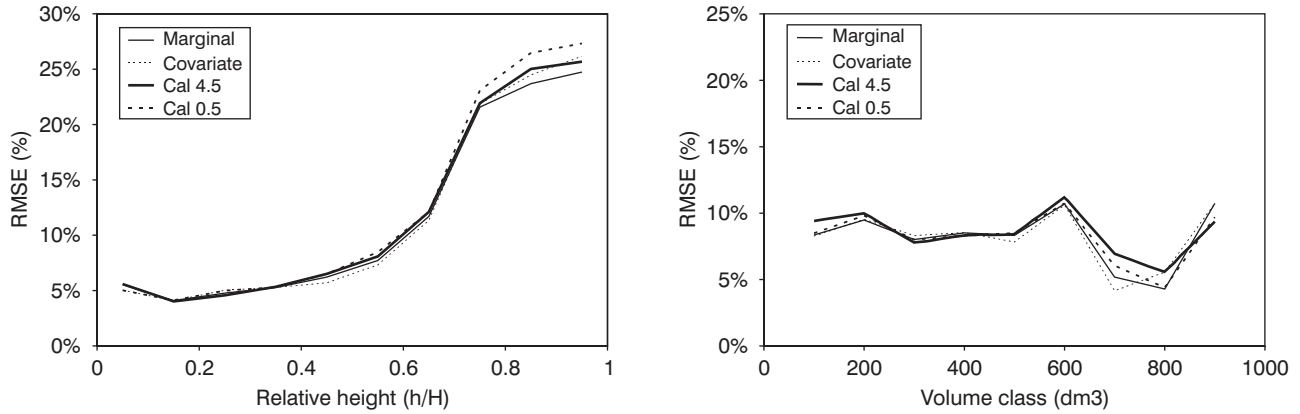


Figure 6. RMSE of relative section diameter (left) and tree volume (right) for the different analysed alternatives. Calibration use data from two additional trees within the plot.

error (bias) and relative error (% bias), as well as its level of significance, for both section diameter and total volume, as a function of relative height and tree volume classes, respectively. These values are compared with the mean value for real and relative errors obtained after calibrating using data from two additional trees measured at 5 and 45 dm (tables 4 and 5).

As shown in these tables, section diameter bias tends to be larger in the upper part of the stem, although they are generally under 5% or 1 cm. Calibrated models tend to give unbiased estimates

($p > 0.01$) for more parts of the stem than the fixed ones, although differences are not very significant. Calibrations using section diameter at 5 dm give unbiased estimates for section diameters up to a relative height of 0.6, while they overestimate section diameter for the upper part of the stem. In evaluating bias in volume prediction, differences between fixed and calibrated alternatives are more obvious. Calibration using 5 dm height measurements give unbiased volume estimates ($p > 0.05$) for all volume classes, except for larger trees (over 900 dm³), while marginal models tend to underestimate tree volume

Table 4. Mean value for real (Bias) and relative error (%Bias) in section diameter prediction. Comparison between marginal, marginal covariate and calibration (using two trees per plot) alternatives. Level of significance is referred to relative error mean value

Relative height class	n	Marginal			Marginal Covariate			Calibration 45 dm			Calibration 5 dm		
		Bias (mm)	Bias (%)	pr > t	Bias (mm)	Bias (%)	pr > t	Bias (mm)	Bias (%)	pr > t	Bias (mm)	Bias (%)	pr > t
0.05	412	7.97	2.49	<0.001	4.65	1.40	<0.001	7.90	2.43	<0.001	5.54	1.70	<0.001
0.1	267	0.82	0.12	0.636	1.22	0.28	0.269	0.47	0.02	0.949	1.24	0.27	0.278
0.2	299	0.14	-0.32	0.241	2.56	0.73	0.012	-0.60	-0.55	0.035	1.20	0.11	0.694
0.3	311	-0.45	-0.72	0.016	2.28	0.67	0.025	-1.39	-1.05	<0.001	-0.35	-0.67	0.025
0.4	304	1.34	-0.02	0.952	3.40	1.23	<0.001	0.46	-0.40	0.287	-0.08	-0.71	0.055
0.5	297	6.20	2.45	<0.001	7.35	3.36	<0.001	5.50	2.08	<0.001	3.44	0.94	0.054
0.6	266	10.68	4.59	<0.001	11.26	5.34	<0.001	9.45	3.84	<0.001	7.42	2.62	<0.001
0.7	251	4.38	-0.74	0.587	4.31	-0.53	0.704	3.28	-1.57	0.255	0.95	-3.43	0.018
0.8	132	-1.45	-5.32	0.009	-2.14	-5.84	0.005	-2.36	-6.20	0.004	-5.08	-8.94	<0.001
0.9	6	6.60	3.49	0.739	4.11	0.78	0.944	4.02	0.73	0.946	1.09	-2.41	0.832

n: number of observations.

Table 5. Mean value for real (Bias) and relative error (%Bias) in tree volume prediction. Comparison between marginal, marginal covariate and calibration (using two trees per plot) alternatives. Level of significance is referred to relative error mean value

Volume Class (dm ³)	n	Marginal			Marginal Covariate			Calibration 45 dm			Calibration 5 dm		
		Bias (dm ³)	Bias (%)	pr > t	Bias (dm ³)	Bias (%)	pr > t	Bias (dm ³)	Bias (%)	pr > t	Bias (dm ³)	Bias (%)	pr > t
100	33	-0.9	-1.40	0.334	1.1	0.42	0.771	-1.7	-2.22	0.164	-1.5	-1.93	0.187
200	35	0.4	0.27	0.867	2.3	1.17	0.469	-0.9	-0.39	0.816	-0.9	-0.37	0.826
300	46	7.1	2.36	0.042	11.3	3.80	<0.001	4.2	1.38	0.218	2.6	0.80	0.498
400	47	13.3	3.37	0.005	16.4	4.15	<0.001	13.5	3.40	0.003	8.0	2.02	0.091
500	24	15.2	3.20	0.052	19.4	4.05	0.006	13.9	2.96	0.074	8.6	1.83	0.289
600	10	38.1	6.60	0.028	38.1	6.59	0.027	33.9	5.88	0.068	34.3	5.95	0.055
700	7	-2.3	-0.25	0.903	-0.9	-0.10	0.953	-2.0	-0.17	0.949	-6.7	-0.86	0.712
800	6	18.5	2.25	0.168	21.6	2.63	0.225	26.9	3.26	0.075	14.1	1.70	0.308
>900	17	99.4	7.99	<0.001	94.8	7.98	<0.001	71.8	5.79	0.003	81.9	6.52	0.001

n: number of observations.

for most of the classes. In any case, bias in volume estimation is generally close to 5%, except for the largest trees.

Discussion and Conclusions

Taper functions and single tree volume equations, which only include total height and breast height diameter of the tree as predictor variables, are usually constructed and applied within a local area. The use of these functions in a wider range of geographical locations is problematic, since they are then applied to trees growing in stands with different ecological or silvicultural conditions, factors which affect stem form and volume (Larson, 1963; Muhairwe *et al.*, 1994; Gregoire and Schabenberger, 1996). In the present work, a method is proposed to deal with stem form differences between trees and between stands by formulating taper functions as multilevel mixed models (Lappi, 1986; Gregoire and Schabenberger, 1996; Tassisa and Burkhart, 1998; Eerikäinen, 2001), including random components at tree and plot level, whilst also taking into account the heterocedastic and autocorrelative pattern for observations coming from the same tree.

Multilevel mixed models enable us to obtain estimates for the variance components determining random variability in stem form at different levels. In our case, greater variability in stem form is detected

between plots, a result similar to that obtained by Lappi (1986), with *Pinus sylvestris* L. On the other hand, Tassisa and Burkhart (1998) found for *Pinus taeda* L. that stem form variability was greater between trees in the same stand than between stands. This difference in results can be explained because Tassisa and Burkhart's model included dbh/H ratio and thinning intensity, covariates which already explained between-stand variability, while the other basic models only included variables associated with relative height and diameter (as is our case) or size (as in Lappi's approach).

A high correlation pattern is detected between adjacent observations coming from the same tree (within tree level). Autocorrelation and heterocedasticity are typical difficulties associated with data used to fit taper equations. The joint use of a variance function and a non-diagonal covariance structure matrix, as proposed by Fang and Bailey (2001), has allowed us to obtain efficient estimates for the parameters in the model. The power covariance structure selected considers that the level of correlation among observations from the same tree is inversely related to their separation along the stem. This structure has been previously identified (Álvarez *et al.*, 2004) as the most appropriate for fitting taper functions.

Several covariates were evaluated to explain random variability in stem form at plot and tree level. Basal area enters the model with a negative parameter, affecting a term (C_{ijk}) whose value is positive for $h > 13$ and negative if $h < 13$. This means that trees growing in

high densities show a greater pattern of narrowing than trees growing in open stands, as stated in fig. 4a. On the other hand, the ratio between breast height diameter and mean square diameter dbh/dg enters the model with a positive parameter, affecting term B_{ijk} , whose value is negative if $h < 13$ and positive if $h > 13$. The inclusion of this social status index means that dominant trees show a more cylindrical stem form than dominated ones (fig. 4b) and, in the case of two trees of the same size in different stands with similar stocking conditions, the one which occupies a dominant position will produce a larger quantity of timber.

Covariates entering the model indicate that trees growing free from competition are less tapered than trees growing under large stocking densities. This result conflicts with both Larson's original theory (1963) and Assmann's study (1970) as well as with recent studies confirming these theories (e.g., Peltola *et al.*, 2002). To explain this behaviour, first we should consider that stand density or social position affect stem form through crown ratio and length, and that stem within crown tends to be more tapered than branch-free bole, due to the contribution of branches to stem growth (as was suggested as early as 1864 by Pressler, or Hartig, 1870; both references in Larson, 1963). In most of the analysed species, trees growing free from competition usually show a large crown ratio and a large-tapered shape. However, free-grown stone pine trees have a polyarchic character and lack of apical dominance, conforming umbrella-like crowns (Lanner, 1989), where foliar biomass is concentrated in the upper part of the stem (a phenomenon increased by stem pruning). This leads to small crown length and ratio, a large, branch-free bole, and a cylindrical shape for most of the stem. On the other hand, in highly stocked stone pine stands, the crown pattern and branching development tend to be similar to other coniferous species (as pointed out by Mutke *et al.* 2005), maintaining a single apical dominance which leads to larger crown ratios.

The high level of non-explained random variability indicates the presence of factors not under consideration controlling stem form which act at plot (e.g. ecological factors or silvicultural treatments) and tree (e.g. microsite, distance dependent competition, crown, genetics...) levels. Future research might focus on identifying the main factors affecting stem form. In an attempt to bridge this information gap, calibration was proposed as an alternative to improve taper

models. Calibration at unsampled units is one of the main advantages of using mixed models, as stated in forest research since Lappi and Bailey's work (1988), or, more recently, in the works by Lappi (1991), Fang and Bailey (2001) or Calama and Montero (2004).

In the present work, we compared two calibration alternatives based on measuring additional stem diameter sections on different trees in the plot at 5 dm or 45 dm height. The use of additional section diameter measurements as explanatory covariates in stem taper or volume equations has been a common practice in forest research (Laasasenaho, 1982; Kilkki, 1983; Martínez-Millán *et al.*, 1992; Kozak, 1998). In these works, it was necessary to measure this additional section for all the trees, while in the proposed calibration approach, tree volume predictions are improved if just one additional measurement is taken for a single tree per plot.

Calibration does not lead to a substantial increment in section diameter predictive accuracy, although calibrated diameter estimations are less biased than marginal ones. Calibrated predictions using 5 dm height measurements are less biased, leading to unbiasedness and significant reductions in error for tree volume predictions. With respect to the number of trees included in the calibration sample, the differences between including three, four or five trees are small, so it is assumed that the inclusion of a larger number of trees will not improve predictions. The evaluation of both RMSE and bias trend on section diameter across different relative heights show that calibration, especially that which uses 5 dm heights, is significant in that it reduces bias, mainly in the lower sections. As a result, calibrated tree volume predictions are more accurate and unbiased than marginal ones, since section diameters are more adequately predicted in the lower part of stem, where a large amount of timber is concentrated. Unbiased and accurate single tree volume prediction is useful, for example, in determining stand volume in management inventories. From this point of view, the use of three additional section diameter measurements at 5 dm height per plot could be interesting to improve stand volume estimates.

The main drawback of the model is the biased, less accurate trend for section diameter estimates in the upper part of the stem. This can be explained by the different pattern of stem narrowing that stone pine trees show when the section is located within the crown, where section diameter decreases quite rapidly from

shoot to shoot. To take into account this pattern it would be necessary to include nonlinear parameters in the model, or at least, define a different stem shape for the part of the stem above the base of the crown (Burkhart and Walton, 1985). In our case, as crown dimensions are not available and the amount and value of the timber in that part of the stem is quite low (mainly dedicated to fuelwood), we decided to maintain the simpler linear structure.

Calibration leads to a slight improvement of the model when compared to the marginal basic and covariate models. The efficiency of the marginal and marginal covariate predictions, without calibration, might then seem sufficient in most cases. Even by using a simple OLS taper function (see table 2, last column), the predictive ability is similar, although parameter estimation is not so efficient. If we take into account that calibration means increasing inventory costs, the proposed model should only be calibrated when very accurate volume estimates are required, or in those cases where complementary data are already available.

In any case, it is important to point out that, despite the scarce results obtained through calibration, the methodology proposed in the present work could be useful for formulating taper functions for other species showing a more regular pattern of stem narrowing. Another possibility for improving predictions would be to use the methodology for fitting a more flexible nonlinear taper function, (e.g., variable exponent taper equation, such as that formulated by Kozak, 1988), but in this case, the covariance structure formulation would be more complex. Finally, both multilevel mixed models as well as calibration should be considered useful techniques for modelling other tree-level variables which display patterns of spatial or temporal correlation.

Acknowledgements

Authors wish to express their gratitude to Dr. Juha Lappi, Dr. Miren del Río and two anonymous reviewers, whose remarks and suggestions over the manuscript improved it significantly. The research was partially supported by a grant to the corresponding author from the Consejería de Educación, Comunidad de Madrid. The research was carried out in the context of the INIA funded project SC-99-017.

References

- ÁLVAREZ J.G., RUIZ A.D., BARRIO M., 2004. Modelos paramétricos de la matriz de varianzas-covarianzas en el ajuste de funciones de perfil. *Cuad. Soc. Esp. Cie. For.* 18, 297-302.
- AMIDON E.L., 1984. A general taper functional form to predict bole volume for five mixed-conifer species in California. *For. Sci.* 30 (1), 166-171.
- ASSMANN E., 1970. The principles of Forest Yield Study. Pergamon Press, Oxford. 506 pp.
- BENNET F.A., SWINDEL B.F., 1972. Taper curves for planted slash pine. USDA Forest Service Research Note 179, 4 pp.
- BONDESON J., 1990. Prediction in random coefficients regression models. *Biometric Journal* 32, 387-405.
- BRUCE D., CURTIS R.O., VANCOEVERING C., 1968. Development of a system of taper and volume tables for red alder. *For. Sci.* 14 (3), 339-350.
- BURKHART H.E., WALTON S.B., 1985. Incorporating crown ratio into taper equation for loblolly pine trees. *For. Sci.* 31 (2), 478-484.
- CALAMA R., CAÑADAS N., MONTERO G., 2003. Inter-regional variability in site index models for even-aged stands of stone pine (*Pinus pinea* L.) in Spain. *Ann. For. Sci.* 60(3): 259-269.
- CALAMA R., MONTERO G., 2004. Interregional non-linear height-diameter model with random coefficients for stone pine in Spain. *Can. J. For. Res* 34: 150-163.
- CASTEDO F., ÁLVAREZ J.G., 2000. Construcción de una tarifa de cubicación con clasificación de productos para *Pinus radiata* D. Don en Galicia basada en una función de perfil de tronco. *Invest. Agr. Sist. Recur. For.* 9 (2), 253-269.
- DAVIDIAN M., GILTINAN D.M., 1993. Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics* 49, 59-73.
- EERIKÄINEN K., 2001. Stem volume models with random coefficients for *Pinus kesya* in Tanzania, Zambia and Zimbabwe. *Can. J. For. Res.* 31, 879-888.
- FANG Z., BAILEY R.L., 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For. Sci.* 47 (3), 287-300.
- FARRAR R.M., 1987. Stem profile functions for predicting multiple product volumes in natural longleaf pineas. *Southern Journal of Applied Forestry* 11 (3), 161-167.
- FOX J.C., ADES P.K., BI H., 2001. Stochastic structure and individual-tree growth models. *For. Ecol. Manage.* 154, 261-276.
- GREGOIRE T.G., 1987. Generalized error structure for forestry yield models. *For. Sci.* 33: 423-444.
- GREGOIRE T.G., SCHABENBERGER O., 1996. A nonlinear mixed-effects model to predict cumulative bole volume of standing trees. *Journal of Applied Statistics* 23, 257-271.
- GREGOIRE T.G., SCHABENBERGER O., BARRETT J.P., 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Can. J. For. Res.* 25, 137-156.

- HARVILLE D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Assoc.* 72, 320-340.
- HENDERSON C.R., KEMPTHORNE O., SEARLE S.R., KROSIGK C.M., 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192-217.
- HOX J.J., 1995. Applied multilevel analysis. TT-Publikaties, Amsterdam. 119 pp.
- ICONA, 1990. Segundo Inventario Forestal Nacional. Explicaciones y métodos. Ministerio de Agricultura, Pesca y Alimentación., Madrid. 174 pp.
- KILKKI P., 1983. Sample trees in timber volume estimation. *Acta Forestalia Fennica* 182, 1-31.
- KOZAK A., 1988. A variable-exponent taper equation. *Can. J. For. Res.* 18, 1363-1368.
- KOZAK A., 1998. Effects of upper stem measurements on the predictive ability of a variable-exponent equation. *Can. J. For. Res.* 28, 1078-1083.
- KOZAK A., MUNRO D.D., SMITH J.G., 1969. Taper functions and their application in forest inventory. *For. Chron.* 45 (4), 278-283.
- LAASASENAHO J., 1982. Taper curve and volume functions for pine, spruce and birch. *Communications Instituti Forestalis Fenniae* 108, 1-74.
- LAIRD N.M., WARE J.H., 1982. Random-effects models for longitudinal data. *Biometrics* 38, 963-974.
- LANNER R.M., 1989. An observation on apical dominance and the umbrella crown of Italian stone pine (*Pinus pinea*, Pinaceae). *Econ. Bot.* 43, 128-130.
- LAPPI J., 1986. Mixed linear models for analyzing and predicting stem form variation of scots pine. *Communicationes Instituti Forestalis Fenniae* 134, 1-69.
- LAPPI J., 1990. Statistical methods for changing and nonhomogeneous forests. XIX IUFRO World Congress. Montreal, Canada, 115-122.
- LAPPI J., 1991. Calibration of height and volume equations with random parameters. *For. Sci.* 37 (3), 781-801.
- LAPPI J., BAILEY R.L., 1988. A height prediction model with random stands and tree parameters: an alternative to traditional site index methods. *For. Sci.* 34(4), 907-927.
- LARSON, P.M., 1963. Stem form development of forest trees. *For. Sci. Monogr.* 5. 42 pp.
- MAPA, 2001. Anuario de Estadística Agroalimentaria 2000. Ministerio de Agricultura, Pesca y Alimentación, Madrid. 699 pp.
- MARTÍNEZ-MILLÁN J., ARA P., GONZÁLEZ I., 1992. Tablas de cubicación de tres entradas para su uso en la Ordenación de Montes. *Invest. Agr. Sist. Recur. For.* 1 (1), 95-102.
- MARTÍNEZ-MILLÁN J., ARA P., GONZÁLEZ I., 1993. Ecuaciones alométricas de tres variables: estimación del volumen, crecimiento y porcentaje de corteza de las principales especies maderables españolas. *Invest. Agr. Sist. Recur. For.* 2 (2), 211-229.
- MONTERO G., ALONSO A., RUÍZ-PEINADO R., CAÑELLAS I., CANDELA J.A., PAVÓN J., 2002. La fijación de CO₂ por las masas de pino piñonero en Andalucía. *Forestalia* 7, 8 pp.
- MUHAIRWE C.K., LE MAY V.M., 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.
- MUTKE S., SIEVÄNEN R., NIKINMAA E., PERTTUNEN J., GIL L. 2005. Crown architecture of grafted stone pine (*Pinus pinea* L.): shoot growth and bud differentiation. *Trees* 19: 15-25.
- NEWNHAM R.M., 1992. Variable-form taper functions for four Alberta species. *Can. J. For. Res.* 22, 210-223.
- NOVO N., ROJO A., ÁLVAREZ J.G., 2003. Funciones de perfil de tronco y tarifas de cubicación con clasificación de productos para *Pinus sylvestris* L. en Galicia. *Invest. Agr. Sist. Recur. For.* 12 (1), 123-136.
- PELTOLA H., MIINA J., ROUVINEN I., KELLOMAKI S., 2002. Effect of early thinning on the diameter growth distribution along the stem of scots pine. *Silva Fennica* 36 (4), 813-825.
- PITA P.A., 1967. Tablas de cubicación por diámetros normales y alturas totales. Instituto Forestal de Investigaciones y Experiencias. Ministerio de Agricultura., Madrid.
- PRIETO A., TOLOSANA E., 1991. Funciones de perfil para la cubicación de árboles en pie con clasificación de productos. *Comunicaciones INIA. Serie: Recursos Naturales* 58, 1-84.
- PRODAN M., PETERS R., COX F., REAL P. 1997. Mensura Forestal. IICA-GTZ. San José (Costa Rica). 561 pp.
- REED D., GREEN E., 1984. Compatible stem taper and volume ratio equations. *For. Sci.* 30 (4), 977-990.
- SEARLE S.R., 1971. Linear models. John Wiley & Sons, New York. 532 pp.
- SEARLE S.R., CASELLA G., MCCULLOCH C.E., 1992. Variance Components. John Wiley, New York. 501 pp.
- SHEU C.F., SUZUKI S., 2001. Meta-Analysis using linear mixed models. *Behavioural Research Methods, Instruments and Computers* 33 (2), 102-107.
- SINGER J.D., 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioural Statistics* 23, 323-355.
- TASSISA G., BURKHART H.E., 1998. An application of mixed effects analysis to modeling thinning effects on stem profile of loblolly pine. *For. Ecol. Manage.* 103, 87-101.
- THOMSON A., BAARCLAY H., 1984. Effects of thinning and urea fertilization on the distribution of area increment along the bole of Douglas fir at Shawnigan Lake. *Can. J. For. Res.* 14, 879-884.
- VALENTINE H.T., GREGOIRE T.G., 2001. A switching model of bole taper. *Can. J. For. Res.* 31, 1400-1409.
- VONESH E.F., CHINCHILLI V.M., 1997. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, Inc., New York. 560 pp.
- WEST P.W., RATKOWSKY D.A., DAVIS A.W., 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. *For. Ecol. Manage.* 7, 207-224.