

Merchantable volume system for pedunculate oak in northwestern Spain

Marcos BARRIO ANTA^{a,d,*}, Ulises DIÉGUEZ-ARANDA^a, Fernando CASTEDO-DORADO^b,
Juan Gabriel ÁLVAREZ GONZÁLEZ^a, Klaus VON GADOW^c

^a Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain

^b Departamento de Ingeniería Agraria, Universidad de León. Escuela Superior y Técnica de Ingeniería Agraria, Avenida de Astorga, 24400 Ponferrada, Spain

^c Institut für Waldinventur und Waldwachstum, Georg-August-Universität Göttingen, Büsingenweg 5, 37077 Göttingen, Germany

^d Current Address: Departamento de Sistemas y Recursos Forestales, CIFOR-INIA. Carretera de la Coruña km. 7, 28080 Madrid, Spain

(Received 23 August 2006; accepted 20 December 2006)

Abstract – A model is required for accurate estimation of the merchantable volume of pedunculate oak (*Quercus robur* L.) trees in Galicia, northwestern Spain. Accordingly, the purpose of the present study was to obtain equations for predicting merchantable volumes and stem profiles of individual trees. For this reason, two compatible and four non-compatible volume systems were initially evaluated and fitted to data from 251 destructively sampled trees which were collected in stands located throughout the area of distribution of the species in Galicia. The outliers were removed to provide a data set of measurements from 3 090 sections, which was then available for fitting. A second-order continuous autoregressive error structure was used to account for autocorrelation. Comparison of the models was carried out using overall goodness-of-fit statistics and box plots of residuals against relative height or diameter class. The compatible volume system of Fang et al. [22] provided the best compromise in describing the stem profile and estimating merchantable height, merchantable volume and total volume and is therefore recommended for pedunculate oak stands in Galicia.

taper function / volume ratio / Galicia / *Quercus robur*

Résumé – Un modèle pour l'estimation du volume commercialisable de chêne pédonculé dans le Nord-Ouest de l'Espagne. Un modèle d'estimation précise du volume marchand du chêne pédonculé (*Quercus robur* L.) à l'échelle des individus en Galice au Nord-Ouest de l'Espagne était indispensable. En conséquence, cette étude a été réalisée pour obtenir des équations de prédiction du volume marchand et du profil de tronc d'arbres individuels. Pour cela, deux modèles compatibles et quatre modèles non compatibles de volume ont été analysés et comparés à un échantillon de 251 arbres récoltés dans des peuplements de la zone de distribution de l'espèce en Galice. Les données extrêmes ont été éliminées pour fournir un ensemble de 3 090 sections disponibles pour l'étude. Une structure d'erreur auto-régressive continue a été utilisée pour prendre en compte l'autocorrélation. La comparaison des modèles a été réalisée en utilisant les statistiques de meilleur ajustement et les nuages de points des résidus comparés aux classes de hauteur relative ou de diamètre. Le système de volume compatible de Fang et al. [22] a fourni le meilleur compromis pour la description du profil de tronc et l'estimation de la hauteur marchande, du volume marchand et du volume total et son usage est donc recommandé pour les peuplements de chêne pédonculé en Galice.

fonction profil / rapport de volume / Galice / *Quercus robur*

1. INTRODUCTION

Determination of stem standing volume is very useful for both research and practical purposes in forestry, and contributes to the sustainable management of timber resources. The currently-used method of measuring the volume of sample trees is expensive, and volume equations are therefore required for predicting individual tree volume by measuring diameter at breast height (D) and total height (H). The volume of interest may comprise the entire bole or only the portion between the stump level and a specific point on the upper bole. Equations that allow determination of the volume to a

minimum diameter or height that establishes a merchantability threshold are known as merchantable volume equations. These equations have commonly been obtained by developing volume ratio equations or taper functions.

Volume ratio equations predict the ratio of the merchantable volume to the total-bole volume of the tree. This approach was introduced by Burkhart [11] and has been widely used (e.g., [13, 16, 26, 43, 54, 57–59]). Basically, the ratio (r) between the cumulated volume at each tree section (v) and the total volume of the tree (V) is related to measured values of D and/or H and a variable that specifies the upper-bole diameter (d) or height (h). If the upper-bole diameter appears as a variable in the model, the volume ratio equation will provide the merchantable volume to that diameter (v); if the variable is the

* Corresponding author: barrio@lugo.usc.es

upper-bole height, the model will provide the merchantable volume to that height (v_h). The two components of the volume ratio (r and V) equations are often developed separately, and the merchantable volume is expressed as the product $V \cdot r$ [1]. However, other authors (e.g. [10,25,35,53]) have fitted volume ratio equations as a composite model that includes both terms (r and V) in the same expression, in which case total volume becomes a special case of the volume ratio equation, i.e., when $h = H$ or $d = 0$, then $V = f(D, H)$.

Taper functions were first introduced by Höjer (1903, cited in [5]) in an attempt to describe the stem profile, and since then they have always been a matter of interest for foresters (e.g., [7, 8, 22, 31, 33, 39, 44]). The development of a volume equation from a taper function is based on its capability to estimate the diameter at a certain height from ground level. Taking into account the application of the definite integral for computing the volume of a solid of rotation, it is possible to determine the volume between any two heights on the stem. Once the diameter limit is specified, its corresponding height can be determined by inverting the taper function analytically or by using an iterative procedure. Therefore, if merchantable volume to a certain diameter is required, two steps are necessary: determination of the height corresponding to a specific diameter limit and then calculation of the volume to that height by integration.

Ideally, a taper equation should be compatible, i.e. the volume computed by integration of the taper function should be equal to that calculated by a total volume equation [16, 18, 22]. The total volume equation is very easy to use and is therefore preferred when classification of the products by merchantable sizes is not required. It is also desirable to achieve compatibility between the volume computed by integration of the taper function from ground to any merchantable height and the volume obtained with a merchantable volume equation. A comprehensive review of compatible volume systems has been reported by Diéguez-Aranda et al. [21].

Pedunculate oak (*Quercus robur* L.) is distributed throughout most of Europe, with its southwestern range limited by the Iberian Peninsula [19]. This oak is also the dominant tree species in native forest in areas of Galicia (northwest Spain) with a wet oceanic climate, where currently 187 000 ha of pure oak stands are found [62]. Because of its importance and wide occurrence, studies related to the autoecology [20], site quality [4], and carbon and nutrient stocks [2] of pedunculate oak in the area have been carried out recently; nevertheless, none of them was related to merchantable volume estimation. The objective of the present study was therefore to develop a merchantable volume system for this species in Galicia. Three different systems are compared which had provided favourable results in previous studies: a volume ratio equation, a non-compatible taper function and a compatible taper function. The intention is to analyze the advantages and disadvantages of each system. The purpose of the analyses is to address the three main problems associated with merchantable volume equation development: the error structure of the data, the multicollinearity among independent variables, and the heteroscedasticity.

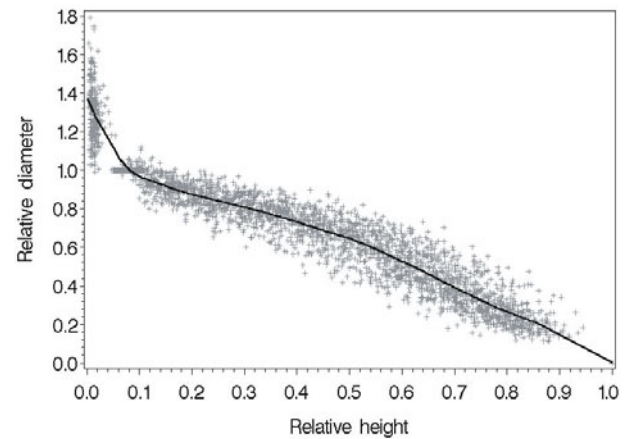


Figure 1. Data points of relative diameter (quotient between an upper-stem diameter and the diameter at breast height) and relative height (quotient between height above ground level to an upper-stem diameter and total tree height) plotted with a local regression *loess* smoothing curve (smoothing factor = 0.25).

2. MATERIAL AND METHODS

2.1. Data set

Data corresponding to 251 non-forked trees were collected in stands located throughout the area of distribution of pedunculate oak in Galicia, covering the existing range of ages, stand densities and site qualities. The trees were selected to ensure a representative distribution by diameter classes. Diameter at breast height (1.3 m above ground level) was measured to the nearest 0.1 cm in each tree. The trees were later felled leaving stumps of average height 0.22 m, and total bole length was measured to the nearest 0.01 m. Each tree was cut into logs at 1 to 3 m intervals of bole length until the point where the well-defined bole ended. The number of observations per tree ranged from 5–19. The average of two perpendicular diameters outside bark, measured to the nearest 0.1 cm in each section, was used. Log volumes were determined using Smalian's formula — except for the top section, which was treated as a cone. Outside bark total stem volume (above stump) was obtained by summing the outside bark log volumes and the volume of the top section of the tree.

The scatter plot of relative diameter against relative height was examined visually to detect possible anomalies in the data caused by wounds, epicormic branches or coarse bark. As taper functions are not intended for deformed stems, abnormally shaped sections were not included in further analysis. For this reason, a nonparametric taper curve was fitted by local regression, with the LOESS procedure of SAS/STAT® [47]. The process was carried out following the procedure described by Bi [7] for detecting abnormal data points. This involved local quadratic fitting with a smoothing parameter of 0.25, which was selected after iterative fitting and visual examination of the smoothed taper curves overlaid on the data. Only 0.32% of the observations in the data set were extreme values. The plot of relative height against relative diameter used for this study, together with the loess regression line, are shown in Figure 1.

Summary statistics including mean, minimum, maximum and standard deviation values of the main tree variables finally used for model fitting are shown in Table I.

Table I. Summary statistics of the 251 oak trees used to develop outside bark taper and volume equations.

Variable	Mean	Minimum	Maximum	S. D.
D (cm)	28.6	8.4	67.5	9.8
H (m)	16.4	8.0	27.6	3.9
V (m ³)	0.57	0.03	3.54	0.51

The following notation will be used hereafter:

D = diameter at breast height outside bark (1.3 m above ground, cm);

d = top diameter outside bark at height h (cm);

H = total tree height (m);

h = height above ground to top diameter d (m);

h_{st} = stump height (m);

V = total stem volume outside bark from stump (m³);

v = merchantable volume outside bark (m³) from stump to the point where diameter = d ;

v_h = merchantable volume outside bark (m³) from stump to the point where height = h ;

a_0 - a_2 , b_0 - b_6 , p_1 , p_2 = coefficients to be estimated;

$k = \pi/40000$, a metric constant for converting from diameter squared in cm² to cross-section area in m²;

$q = h/H$;

$z = (H - h)H$.

2.2. Equations tested

In this study two volume ratio equations, two non-compatible taper functions and two compatible taper functions were first considered for analysis. Within the first group, we chose the version modified by Cao et al. [13] of the model originally proposed by Burkhart [11], and the model derived by Clark and Thomas [15] as a modification of the exponential ratio equation of Van Deusen et al. [58] by adding a third parameter.

Both of these models have been used recently for predicting green weight of loblolly pine trees in the United States [10, 52].

Within the non-compatible taper functions we tested the variable-exponent models developed by Bi [7] and Kozak [33]. These types of functions describe the stem shape with a changing exponent or variable from ground to top of the tree, and are assumed to be the most flexible in their ability to fit data for species and trees with different stem forms, providing the lowest degree of local bias and greatest precision in taper predictions (e.g., [31, 33, 36, 39]).

Within the compatible taper functions we tested the single-polynomial model developed by Goulding and Murray [24] and the exponential-segmented model developed by Fang et al. [22]. These systems provide compatibility between the volume computed by integration of the taper function from ground to top, or any merchantable height, and the volume obtained with a total or merchantable volume equation, respectively. The models selected performed favourably for different species in the studies of Castedo [14], Diéguez-Aranda et al. [21] and Corral et al. [17].

Although all these models fitted the data reasonably well in the preliminary analysis, we finally selected one model from each group for further analysis: Tasissa et al. [52], Kozak [33] and Fang et al. [22].

2.2.1. Tasissa et al. (1997)

This volume system comprises a total volume equation and a volume ratio equation. We used Spurr's combined-variable equation $V = a_0 + a_1 D^2 H$ [50], which performed well in a previous study involving estimation of pedunculate oak tree volume in Galicia [3]. The composite model for predicting merchantable volume to any upper diameter limit is:

$$v = (a_0 + a_1 D^2 H) e^{(b_0 d^{b_1} / D^{b_2})}. \quad (1)$$

The exponential ratio model is doubly constrained because as the upper diameter limit increases from zero to infinity, the ratio decreases to zero, and if the upper diameter limit decreases to 0, the ratio increases to one, and merchantable total volume (v) is equal to total volume (V) [10].

The volume up to any specific upper height limit can also be predicted with the following expression:

$$v_h = (a_0 + a_1 D^2 H) e^{(b_3 (H-h)^{b_4} / H^{b_5})}. \quad (2)$$

The exponential ratio model is constrained in the same way as Equation (1) so that as the upper height limit increases from zero to H the ratio increases to one, and merchantable total volume (v_h) is equal to total volume (V) [10].

Since both values of the volume ratio for any specified diameter limit or its corresponding height must be the same, by equalling and rearranging Equations (1) and (2), two taper functions can be derived to predict diameter outside bark along the stem (d) and the height at which that diameter occurs (h):

$$d = (b_3/b_0)^{1/b_1} D^{b_2/b_1} \left((H-h)^{b_4/b_1} / H^{b_5/b_1} \right) \quad (3)$$

$$h = H - (b_0/b_3)^{1/b_4} (H)^{b_5/b_4} \left(d^{b_1/b_4} / D^{b_2/b_4} \right). \quad (4)$$

2.2.2. Kozak (2004)

This is a variable-exponent taper model, proposed by Kozak [33] as the best solution for consistently estimating the diameter along the stem, the tree or log volume and the merchantable height. The expression of this non-compatible taper function is:

$$d = a_0 D^{a_1} H^{a_2} x^{b_1 q^4 + b_2 (1/e^{D/H}) + b_3 x^{0.1} + b_4 (1/D) + b_5 H^w + b_6 x} \quad (5)$$

$$\text{where } x = \frac{w}{1 - (1.3/H)^{1/3}} \quad \text{and} \quad w = 1 - q^{1/3}.$$

2.2.3. Fang et al. (2000)

This exponential-segmented-compatible model assumes that the stem has three sections with a variable-form factor, although constant for each one. The system comprises three equations: a taper function, a total volume equation and a merchantable volume equation. The expression of the system is:

Taper function:

$$d = c_1 \sqrt{H^{(k-b_1)/b_1} (1-q)^{(k-\beta)/\beta} \alpha_1^{I_1+I_2} \alpha_2^{I_2}} \quad (6)$$

where

$$\begin{cases} I_1 = 1 & \text{if } p_1 \leq q \leq p_2; 0 \text{ otherwise} \\ I_2 = 1 & \text{if } p_2 \leq q \leq 1; 0 \text{ otherwise} \end{cases}$$

p_1 and p_2 are relative heights from ground level where the two inflection points assumed in the model occur

$$\begin{aligned}\beta &= b_1^{1-(I_1+I_2)} b_2^{I_1} b_3^{I_2} & \alpha_1 &= (1-p_1)^{\frac{(b_2-b_1)k}{b_1 b_2}} & \alpha_2 &= (1-p_2)^{\frac{(b_3-b_2)k}{b_2 b_3}} \\ r_0 &= (1-h_{st}/H)^{k/b_1} \\ r_1 &= (1-p_1)^{k/b_1} & r_2 &= (1-p_2)^{k/b_2} \\ c_1 &= \sqrt{\frac{a_0 D^{a_1} H^{a_2-k/b_1}}{b_1(r_0-r_1) + b_2(r_1-\alpha_1 r_2) + b_3 \alpha_1 r_2}}.\end{aligned}$$

Merchantable volume equation:

$$\begin{aligned}v &= c_1^2 H^{k/b_1} (b_1 r_0 + (I_1 + I_2)(b_2 - b_1) r_1 \\ &\quad + I_2(b_3 - b_2) \alpha_1 r_2 - \beta(1-q)^{k/\beta} \alpha_1^{I_1+I_2} \alpha_2^{I_2}).\end{aligned}\quad (7)$$

Volume equation:

$$V = a_0 D^{a_1} H^{a_2}.\quad (8)$$

Although the development of the compatible system of Fang et al. [22] is based on total volume Equation (8), any other total volume equation can be used as input into the system by changing the expression of c_1 in Equation (6).

2.3. Model fitting

In the volume ratio system of Tasissa et al. [52], Equations (1) and (2) can be fitted simultaneously. This procedure optimizes volume predictions to a certain height or diameter limit rather than providing a stem profile description (Eqs. (3) and (4)). Equations (1)–(4) can also be considered as a system of simultaneous equations, in which v , v_h , d and h are variables whose values are determined by the system (endogenous variables) and D and H are exogenous, i.e., they are independent variables that do not depend on any of the endogenous variables in the system.

There are three possible ways of fitting the system of Fang et al. [22]: (i) estimation of the parameters of the total volume equation (8) using the total volume observations, substitution of the estimated parameters in the system, and fitting of the remaining parameters; (ii) estimation of the parameters of the taper equation (6) and substitution in the volume functions, or (iii) estimation of all the parameters of the system simultaneously. We adopted the latter approach for both models [22, 52] to optimize the total sum of squared errors of the system, i.e., to minimize both, the diameter at different heights and the volume prediction errors.

For fitting the system of Tasissa et al. [52], the Full Information Maximum Likelihood (FIML) technique was applied to fit all the equations simultaneously with the MODEL procedure of SAS/ETS® [48]. The equations of the system of Fang et al. [22] are not really simultaneous (none of the exogenous variables in one equation of the system appears as dependent on the left-hand side of the other equation; see [30, 45, 63]); however, since the error components of the variables on the left-hand side and right-hand side are significantly correlated at the 5% level ($Corr(v, d) = 0.1934$), simultaneous fitting using the *seemingly unrelated regression* (SUR) technique [63] was applied. This was accomplished using the MODEL procedure of SAS/ETS® [48].

The non-compatible taper function of Kozak [33] was fitted by means of nonlinear regression with the Marquardt algorithm of the SAS/ETS® MODEL procedure [48].

There are several problems associated with stem taper and volume equation analysis that violate the fundamental least squares assumption of independence and equal distribution of errors with zero mean and constant variance; multicollinearity, autocorrelation and heteroscedasticity are three of the most important of these problems. Although the least squares estimates of regression coefficients remain unbiased and consistent under the presence of multicollinearity, autocorrelation and heteroscedasticity, they are not necessarily the most efficient [37]. These problems may seriously affect the standard errors of the coefficients, invalidating statistical tests that include t or F distributions and confidence intervals [38, 42]. Appropriate statistical procedures should therefore be used in model fitting to avoid problems of heteroscedasticity and autocorrelated errors, and models with low multicollinearity should be selected whenever possible [32].

Multicollinearity is a problem associated with almost all forestry data [32]. This refers to the existence of high intercorrelations among the independent variables in multiple linear or nonlinear regression analysis, because some variables represent or measure similar phenomena. We used the condition number to evaluate the presence of multicollinearity among variables in the models analyzed, which is defined as the square root of the ratio of the largest to the smallest eigenvalue of the $\mathbf{W}'\mathbf{W}$ matrix, where \mathbf{W} is the matrix of the partial derivatives of the parameters. According to Belsey [6], if the condition number is 5–10, multicollinearity is not a major problem, if it is in the range of 30–100, then there are problems, and if it is in the range of 1000–3000 there are severe problems associated with multicollinearity.

In the development of volume ratio equations or taper functions we are dealing with multiple observations on each tree. It is therefore expected that observations within each tree are spatially correlated, which violates the assumption of independent errors. To overcome the possible autocorrelation, we modelled the error term using a continuous autoregressive error structure (CAR(x)), which allows the model to be applied to irregularly spaced, unbalanced data [25, 64]. To account for k -order autocorrelation, a CAR(x) model form that expands the error term in the following way can be used [64]:

$$e_{ij} = \sum_{k=1}^{k=x} I_k \rho_k^{h_{ij}-h_{ij-k}} e_{ij-k} + \varepsilon_{ij}\quad (9)$$

where e_{ij} is the j th ordinary residual on the i th individual, e_{ij-k} is the j -kth ordinary residual on the i th individual, $I_k = 1$ for $j > k$ and it is zero for $j \leq k$, ρ_k is the k -order autoregressive parameter to be estimated, and $h_{ij}-h_{ij-k}$ is the distance separating the j th from the j th- k th observations within each tree, $h_{ij} > h_{ij-k}$. In such cases ε_{ij} now includes the error term under conditions of independence. The above error structure was fitted simultaneously with the mean structure using the MODEL procedure of SAS/ETS® [48], which allows for dynamic updating of the residuals.

Forest modellers are often faced with the problem of heteroscedasticity in their data, especially in the construction of volume equations. Although several assumptions about the nature of the heteroscedasticity problem in the construction of volume equations that depend on two variables were suggested, it is often assumed that the variance of the error (σ_i^2) can be modelled as a power function of $D^2 H$ [12, 23], i.e. $\sigma_i^2 = (D^2 H)_i^m$. The most reasonable value of the exponential term m should provide the most homogeneous studentized residual plot [29]. The value can be obtained by iteratively testing different values of m (e.g., from 0.1 to 2), or optimizing the value using the method suggested by Harvey [27], which consists of using the estimated errors of the unweighted model ($\hat{\varepsilon}_i$) as the dependent variable

in the error variance model, i.e.:

$$\hat{\epsilon}_i^2 = \gamma(D^2H)_i^m \quad (10)$$

or, taking the natural logarithm \ln of the function:

$$\ln \hat{\epsilon}_i^2 = \ln \gamma + m \ln(D^2H)_i \quad (11)$$

The m parameter of Equation (11) was estimated using linear regression. Parameters of Equations (1), (2) and (7) were estimated using generalized non-linear least squares by specifying the weight function with the option $\text{resid.y} = \text{resid.y} / \sqrt{(D^2H)_i^m}$ of the SAS/ETS[®] MODEL procedure [48], where y is the dependent variable of the model.

2.4. Model comparison and model selection

The comparison of the estimates for the different models was based on numerical and graphical analyses of the residuals. Two statistical criteria were examined: the root mean square error (RMSE) and the coefficient of determination for nonlinear regression (R^2). Although there are several shortcomings associated with the use of R^2 in nonlinear regression, the general usefulness of some global measure of model adequacy appears to override some of those limitations [46]. The expressions of these statistics are:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p}} \quad (12)$$

$$R^2 = r_{y_i \hat{y}_i}^2 \quad (13)$$

where y_i and \hat{y}_i are the observed and predicted values of the dependent variable, respectively, n the total number of observations, p the number of model parameters (the number of shared parameters in systems composed of two equations is divided by two), and $r_{y_i \hat{y}_i}$ the correlation coefficient for a linear regression between the observed and predicted values of the dependent variable (see [46], p. 419 and 424).

Although single indices of overall prediction (RMSE and R^2) are good indicators of the effectiveness of each volume system, they may not show the best model for practical purposes. We therefore used box plots for evaluating the distribution of the residuals of the following dependent variables: top diameter (d), height above ground to top diameter (h), merchantable volume (v or v_h) and total volume (V). The first was evaluated by relative height (h/H) classes from ground to top, the second and third by d classes from ground to top, and the fourth by D classes. These graphs are very important for illustrating areas across levels of a grouping variable for which the volume systems provide especially poor or good predictions.

Prior to carrying out this comparison, for the models of Kozak [33] and Fang et al. [22], estimates of the height at which each commercial diameter occurs were obtained. For this purpose, we used the Newton iterative procedure implemented by the SOLVE statement of the SAS/ETS[®] MODEL procedure [48]. In addition, total and commercial tree volumes for the model of Kozak [33] were computed by means of the QUAD subroutine of the SAS statistical package, which performs numerical integration of scalar function in one dimension over finite interval [49].

If we are interested in comparing candidate models in terms of their prediction capabilities, we must take into account that ordinary

residuals are measures of quality of fit and does not assess quality of future prediction ([37], p. 168).

For this purpose, validation of the model must be carried out. Ideally, a newly collected data set will be available for validation [9, 34, 41, 60]. Due to the scarcity of such data, several methods have been proposed (e.g. splitting the data set or cross-validation, double cross-validation), although they seldom provide any additional information compared with the respective statistics obtained directly from models built from entire data sets [34]. Moreover, according to [28, 37] the final estimation of the model parameters should come from the entire data set because the estimates obtained with this approach will be more precise than those obtained from the model fitted from only one portion of the data. Considering these arguments, we decided not to attempt a model validation in this study (relying on the fact that a well developed model will behave well), and to wait for a new data set to evaluate the quality of the predictions.

3. RESULTS AND DISCUSSION

The models were initially fitted without expanding the error term to account for autocorrelation. As expected, because of the hierarchical nature of the data, a trend was apparent in all the models analyzed in terms of the residuals of the taper and merchantable volume equations as a function of lag1- and lag2-residuals within the same tree. We therefore expanded the error term as a continuous autoregressive process, and found that a second order autoregressive structure was sufficient to eliminate the autocorrelation of d predictions in the three models. As an example, Figure 2 (first row) shows how the spatial autocorrelation is removed in the model of Fang et al. [22] after modelling the residuals as a continuous second-order autoregressive structure. Autocorrelation was more severe in merchantable volume than in diameter at a given height, and after applying a third-order autoregressive structure this was only partially removed. This effect is also shown for the merchantable volume model of Fang et al. [22] (Fig. 2, second row). Although accounting for autocorrelation does not improve the predictive ability of the model, it prevents underestimation of the covariance matrix of the parameters, thereby making it possible to carry out the normal statistical hypothesis tests [61]. Therefore, considering the very high significance of all the parameters, including the autocorrelation parameters (Tab. II), it may be assumed that no changes in parameter significance would be found by incorporating the residual's trends in a more efficient way into the model.

For practical purposes, the correlation parameters (Tab. II) can therefore be ignored, unless one is working with several measurements of diameter at different heights on the same individual.

Scatter plots of residuals against predicted merchantable volume showed an increasing variance in residuals as the diameter values increased in Equations (1), (2) and (7). Fitting of Equation (11) by ordinary least squares provided m values of 1.80, 1.71 and 1.52 for inclusion in the weighting factor for the correction for heteroscedasticity in the three merchantable volume equations. These values allowed minimum variance estimates and reliable prediction intervals to be obtained [40].

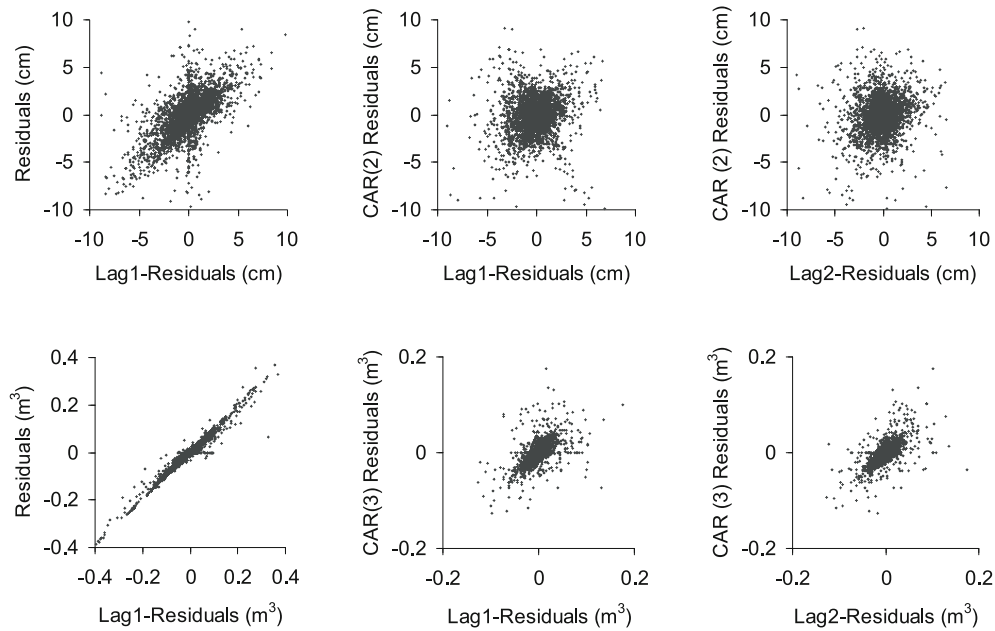


Figure 2. Residuals plotted against lag1-residuals (left column) for Equations (6) and (7) fitted without considering the autocorrelation parameters; residuals lotted against lag1-residuals and lag2-residuals (middle and right columns, respectively) considering a continuous autoregressive error structure of second and third-order, for Equations (6) and (7), respectively.

Table II. Parameter estimates (approximated standard errors in brackets) for the volume systems analyzed.

Parameter	Volume system								
	Tasissa et al. (1997)			Kozak (2004)			Fang et al. (2000)		
	Eq.	Estimate	Approx. std. error	Eq.	Estimate	Approx. std. error	Eq.	Estimate	Approx. std. error
a_0	(1) (2)	0.020512	(0.00280)	(5)	1.055081	(0.0446)	(7)	0.000034	(1.127E-6)
a_1	(1) (2)	0.000003	(1.267E-7)	(5)	0.92845	(0.0121)	(7)	1.896344	(0.0105)
a_2	–	–	–	(5)	0.076525	(0.0197)	(7)	1.127181	(0.0149)
b_0	(1)	–0.30703	(0.0217)	–	–	–	–	–	–
b_1	(1)	3.857896	(0.0448)	(5)	0.681231	(0.0378)	(6) (7)	7.737E-6	(2.229E-7)
b_2	(1)	3.441143	(0.0452)	(5)	–1.03465	(0.1093)	(6) (7)	0.000003	(1.547E-7)
b_3	(2)	–1.10011	(0.0871)	(5)	0.430709	(0.0190)	(6) (7)	0.000041	(5.204E-6)
b_4	(2)	3.52398	(0.0390)	(5)	5.353225	(0.7286)	–	–	–
b_5	(2)	3.318094	(0.0464)	(5)	0.074046	(0.00631)	–	–	–
b_6	–	–	–	(5)	–0.26478	(0.0482)	–	–	–
p_1	–	–	–	–	–	–	(6) (7)	0.061802	(0.00179)
p_2	–	–	–	–	–	–	(6) (7)	0.912443	(0.0446)
ρ_1	(1)	0.969466	(0.00592)	(5)	0.539112	(0.0176)	(6)	0.50109	(0.0175)
ρ_2	(1)	0.791859	(0.00695)	(5)	0.588838	(0.0171)	(6)	0.598486	(0.0163)
ρ_1	(2)	0.791859	(0.00826)	–	–	–	(7)	1.13525	(0.00408)
ρ_2	(2)	0.797918	(0.00867)	–	–	–	(7)	0.963521	(0.00403)
ρ_3	–	–	–	–	–	–	(7)	0.961096	(0.00388)
ρ_1	(3)	0.851988	(0.00509)	–	–	–	–	–	–
ρ_2	(3)	0.753847	(0.00713)	–	–	–	–	–	–
ρ_1	(4)	0.68431	(0.00821)	–	–	–	–	–	–
ρ_2	(4)	0.685797	(0.00851)	–	–	–	–	–	–

Table III. Goodness-of-fit statistics and condition number of the volume systems.

Volume system	Eq.	RMSE	R^2	Condition number
Tasissa et al. (1997)	(1)	0.0654	0.9783	103
	(2)	0.0303	0.9954	
	(3)	1.6126	0.9755	
	(4)	0.8339	0.9745	
Kozak (2004)	(5)	1.9504	0.9725	67
Fang et al. (2000)	(6)	0.0659	0.9779	53
	(7)	1.9955	0.9712	

The parameter estimates and their approximated standard errors for each volume system, after taking into account autocorrelation and heterocedasticity, are shown in Table II. All the parameters were significant at the 5% level.

The values of R^2 and the RMSE are shown in Table III. All the equations performed reasonably well, explaining more than 97% of the total variance of the dependent variables. The highest multicollinearity was found in the system of Tasissa et al. [52], as inferred from the condition number.

The following considerations for each model analyzed can be made on the basis of the goodness-of-fit statistics (Tab. III) and the analysis of the box plots of the residuals for the four dependent variables of interest (d , h , V and v) by relative height or diameter classes (Fig. 3):

- Assessment of the models, taking into account how they perform in describing diameters along the stem, revealed that the variable-exponent model of Kozak [33] performed best and the model of Tasissa et al. [52] performed worst. These results are reasonable since the values of the estimated parameters are optimized for d in the taper function, and for d , h , v and v_h simultaneously, in the system of Tasissa et al. [52]. This model provided a clear underestimation of diameter in the section closest to the ground, whereas the models of Fang et al. [22] and Kozak [33] provided only a very slight overestimation (Fig. 3, first row). Accurate predictions of diameter of this section are important since the basal log is particularly important from a commercial point of view. None of the models of Fang et al. [22] and Kozak [33] showed a systematic trend in the residuals in relation to this variable; however the latter performed slightly better than the former.

- Differences among models as regards the estimations of height along the stem were greater than those corresponding to diameter. Once again, the model of Tasissa et al. [52] underestimated commercial heights by diameter classes and provided the widest interquartile ranges of the residuals (Fig. 3, second row). The other two models provided similar results, with the model of Kozak [33] performing slightly better than the others.

- The predictions of total volume were less accurate for the largest trees (Fig. 3, third row) in all the models analyzed. These results suggest that modelling of butt sweep, which is a characteristic attribute of large oak trees, should be carried

out with more measurements of diameter at different heights in the lower part of the stem (e.g., [33, 51, 55, 56]). The volume systems of Fang et al. [22] and Tasissa et al. [52] performed better than the model of Kozak [33], which clearly underestimated the volumes of the largest trees. Again, the results were consistent with the fact that parameter estimates of the latter model were optimized for predicting stem profile rather than stem volume.

- All the models provided biased merchantable volume estimates in the bole sections closest to the ground level. The system of Tasissa et al. [52] overestimated the merchantable volume, whereas the systems of Fang et al. [22] and Kozak [33] underestimated it, but to a much lesser extent (Fig. 3, last row). The residuals for merchantable volume for the system of Tasissa et al. [52] were also more widely dispersed because this model presented wider interquartile ranges for all the limit diameter classes (Fig. 3). Although the parameters of the model of Kozak [33] are optimized for diameter predictions along the entire tree bole, this model provided good predictions in merchantable volume estimations (obtained by numerical integration, which is somewhat a disadvantage of this model), which were surprisingly better than those provided by the system of Tasissa et al. [52]. With this system, merchantable volume comprising volume ratio-based systems is estimated with more accuracy to any specific top limit height than to any specific top limit diameter, because the RMSE of Equation (1) is approximately one-half the size of the RMSE of Equation (2) (Tab. III). These results are also consistent with those reported by other authors (e.g., [43, 53, 59]).

The above considerations can be summarized as follows for our database: (i) if the main interest of the research is stem profile description, the variable-exponent model proposed by Kozak [33] seems the best option; (ii) if the purpose is to predict total and merchantable volume, the compatible volume system of Fang et al. [22] appears to be the best solution for both variables; and (iii) if a good compromise between stem profile description and total or merchantable volume estimation is required, the compatible volume system of Fang et al. [22] again provides the best results.

Eventually, it is up to the practitioner to decide which is the most appropriate model for the required use. However, considering the good results provided in describing the stem profile and predicting stem volume, we recommend the compatible volume system developed by Fang et al. [22] for general use. Furthermore, this model showed the lowest value for the condition number and therefore minimized the problems associated with multicollinearity (Tab. III).

4. CONCLUSIONS

The present study was carried out on the basis of up-to-date knowledge available in the literature on taper and volume modelling. Among the three different types of models for describing the taper and estimating stem volume for pedunculate oak that were analyzed, the segmented compatible model of Fang et al. [22] best described the experimental data.

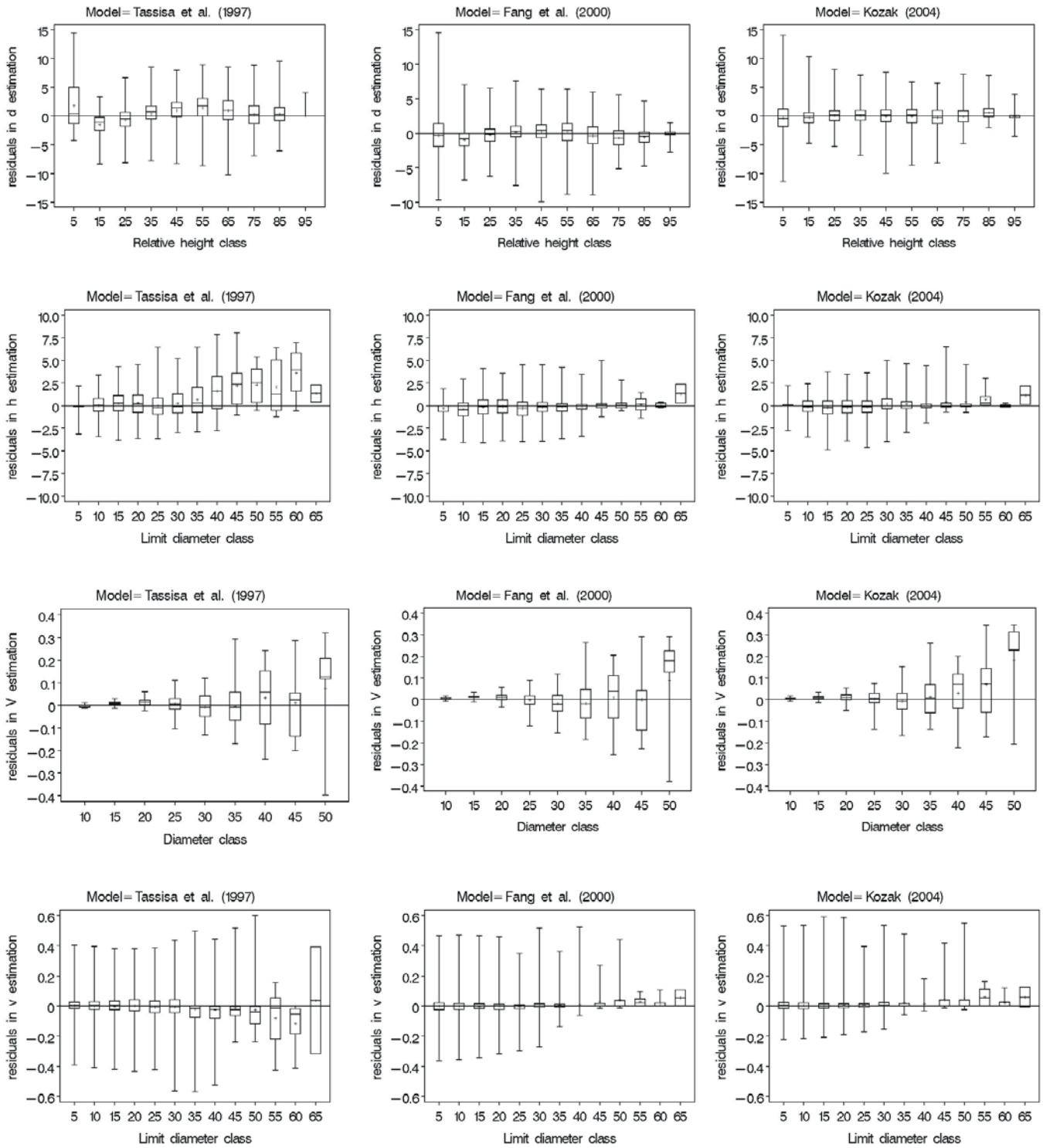


Figure 3. Box plots of: d residuals (cm) against relative height classes (cm), h residuals (m) against relative diameter classes (cm), V residuals (m^3) against diameter at breast height classes (cm) and v residuals (m^3) against limit diameter classes (cm) for the three different models considered. Plus signs represent the mean estimates. Boxes represent the interquartile range. The maximum and minimum height estimates are represented by the upper and lower small horizontal lines crossing the vertical bars, respectively.

This model compared favourably with the noncompatible taper function of Kozak [33] and with the volume-ratio system of Tasissa [52] in its ability to predict upper-stem diameter, height to any specific diameter, total volume, and merchantable volume. The system of Fang et al. [22] also has the advantage that the resulting taper, merchantable and total volume equations are compatible with each other. The satisfactory results obtained with this system in estimating total tree volume are probably due to the fact that the simultaneous model fitting reduced the total system squared error, i.e., it simultaneously minimized both diameter at different heights and volume prediction errors.

The diameter and the cumulative volume along stems were found to be highly correlated. Within-tree diameter autocorrelation was satisfactorily modelled as a second-order continuous autoregressive process, whereas cumulative volume was modelled as second or third-order continuous autoregressive processes. Non constant variance in volume equations was removed by using a power function of the combined variable squared diameter at breast height and total height. Only trees with a clearly defined stem from base to tip and non-forked trees were considered in the present study. To complete the information provided by the system, a new sample considering forked trees of different heights or trees where the well-defined trunk ends at the base of the crown is required.

Acknowledgements: The collection of data for the present study was financially supported by the local government (Xunta de Galicia, project PGIDT99MA29101) and INIA (Spanish National Institute of Agricultural Research). We also thank "CERNA Ingeniería" for his helpful in data collection. Dr. Christine Francis corrected the English grammar of the text.

REFERENCES

- [1] Avery T.E., Burkhart H.E., Forest measurements, 5th ed., McGraw-Hill, New York, 2002.
- [2] Balboa-Murias M.A., Rojo A., Álvarez J.G., Merino A., Carbon and nutrient stocks in mature *Quercus robur* L. stands in NW Spain, *Ann. For. Sci.* 63 (2006) 557–565.
- [3] Barrio M., Álvarez González J.G., Díaz-Maroto I.J., Elaboración de una tarifa con clasificación de productos para *Quercus robur* L. en Galicia basada en un modelo de volumen porcentual, *Invest. Agrar.: Sist. Recur. For.* 13 (2004) 506–517.
- [4] Barrio M., Diéguez-Aranda U., Site quality of pedunculate oak (*Quercus robur* L.) stands in Galicia (northwest Spain), *Eur. J. For. Res.* 124 (2005) 19–28.
- [5] Behre C.E., Preliminary notes on studies of tree form, *J. For.* 21 (1923) 507–511.
- [6] Belsey D.A., Conditioning diagnostics, collinearity and weak data in regression, John Wiley & Sons, Inc., New York, 1991.
- [7] Bi H., Trigonometric variable-form taper equations for Australian eucalyptus, *For. Sci.* 46 (2000) 397–409.
- [8] Brink C., Gadow K.v., On the use of growth and decay functions for modelling stem profiles, *EDV in Medizin und Biologie* 17 (1985) 20–27.
- [9] Bruce D., Wensel L.C., Modelling forest growth: approaches, definitions and problems, in: Ek A.R., Shifley S.R., Burke T.E. (Eds.), *Forest Growth Modelling and Prediction Conference*, USDA For. Ser. Gen. Tech. Rep. NC-120, 1988, p. 1–8.
- [10] Bullock B.P., Burkhart H.E., Equations for predicting green weight of loblolly pine trees in the south, *South. J. Appl. For.* 27 (2003) 153–159.
- [11] Burkhart H.E., Cubic-foot volume of loblolly pine to any merchantable top limit, *South. J. Appl. For.* 1 (1977) 7–9.
- [12] Cailliez F., Estimación del volumen forestal y predicción del rendimiento, FAO, Roma, 1980.
- [13] Cao Q.V., Burkhart H.E., Max T.A., Evaluations of two methods for cubicfoot volume prediction of loblolly pine to any merchantable limit, *For. Sci.* 26 (1980) 71–80.
- [14] Castedo F., Modelo dinámico de crecimiento para las masas de *Pinus radiata* D. Don en Galicia. Simulación de alternativas selvícolas con inclusión del riesgo de incendio, Doctoral Thesis, Escola Politécnica Superior, University of Santiago de Compostela, 2004.
- [15] Clark A., Thomas C.E., Weight equations for southern tree species. When are and what is needed, in: Daniels R.F., Dunhan P.H. (Eds.), *Proceedings of the 1983 southern forest biomass workshop*, USDA Forest Service, Southern Forest Experimental Station, 1984, pp. 100–106.
- [16] Clutter J.L., Development of taper functions from variable-top merchantable volume equations, *For. Sci.* 26 (1980) 117–120.
- [17] Corral-Rivas J., Diéguez-Aranda U., Castedo F., Corral S., A merchantable volume system for major pine species in El Salto, Durango (Mexico), *For. Ecol. Manage.* 238 (2007) 118–129.
- [18] Demaerschalk J., Converting volume equations to compatible taper equations, *For. Sci.* 18 (1972) 241–245.
- [19] Díaz-Fernández P.M., Jiménez P., Martín S., De Tuero Y., Reyna M., Gil L., Regiones de procedencia de *Quercus robur* L., *Quercus petraea* (Matt.) Liebl. y *Quercus humillilis* (Miller), MAPA, Madrid, 1995.
- [20] Díaz-Maroto I.J., Vila-Lameiro P., Silva-Pando F.J., Autoécologie des chaînes de *Quercus robur* L. en Galicie (Espagne), *Ann. For. Sci.* 62 (2005) 737–749.
- [21] Diéguez-Aranda U., Castedo-Dorado F., Álvarez-González J.G., Rojo A., Compatible taper function for Scots pine (*Pinus sylvestris* L.) plantations in north-western Spain, *Can. J. For. Res.* 36 (2006) 1190–1205.
- [22] Fang Z., Borders B.E., Bailey R.L., Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors, *For. Sci.* 46 (2000) 1–12.
- [23] Furnival G.M., An index for comparing equations used in constructing volume tables, *For. Sci.* 7 (1961) 337–341.
- [24] Goulding C.J., Murray J., Polynomial taper equations that are compatible with tree volume equations, *N. Z. J. For. Sci.* 5 (1976) 313–322.
- [25] Gregoire T.G., Schabenberger O., Barrett J.P., Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements, *Can. J. For. Res.* 25 (1995) 137–156.
- [26] Gregoire T.G., Schabenberger O., A non-linear mixed-effects model to predict cumulative bole volume of standing trees, *J. Appl. Stat.* 23 (1996) 257–271.
- [27] Harvey A.C., Estimating regression models with multiplicative heteroscedasticity, *Econometrica* 44 (1976) 461–465.
- [28] Hirsch R.P., Validation samples, *Biometrics*, 47 (1991) 1193–1194.
- [29] Huang S., Yang Y., Wang Y., A critical look at procedures for validating growth and yield models, in: Amaro A., Reed D., Soares P. (Eds.), *Modelling forest systems*, CAB International, Wallingford, Oxfordshire, UK, 2003, 271–293.
- [30] Judge G.G., Carter R., Griffiths, W.E., Lutkepohl H., Lee T.C., *Introduction to the theory and practice of econometrics*, John Wiley & Sons, New York, 1988.
- [31] Kozak A., A variable-exponent taper equation, *Can. J. For. Res.* 18 (1988) 1363–1368.

- [32] Kozak A., Effects of multicollinearity and autocorrelation on the variable exponent taper functions, *Can. J. For. Res.* 27 (1997) 619–629.
- [33] Kozak A., My last words on taper equations, *For. Chron.* 80 (2004) 507–515.
- [34] Kozak A., Kozak R.A., Does cross validation provide additional information in the evaluation of regression models? *Can. J. For. Res.* 33 (2003) 976–987.
- [35] McTague J.P., Bailey R.L., Simultaneous total and merchantable volume equations and a compatible taper function for loblolly pine, *Can. J. For. Res.* 17 (1987) 87–92.
- [36] Muirhead C.K., Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia, *For. Ecol. Manage.* 113 (1999) 251–269.
- [37] Myers R.H., Classical and modern regression with applications, 2nd ed., Duxbury Press, Belmont, California, 1990.
- [38] Neter J., Kutner M.H., Nachtsheim C.J., Wasserman W., Applied linear statistical models, 4th ed., McGraw-Hill, New York, 1996.
- [39] Newnham R., Variable-form taper functions for four Alberta tree species, *Can. J. For. Res.* 22 (1992) 210–223.
- [40] Parresol B.R., Additivity of nonlinear biomass equations, *Can. J. For. Res.* 31 (2001) 865–878.
- [41] Pretzsch H., Biber P., Durský J., Gadow K.v., Hasenauer H., Kändler G., Kenk G., Kublin E., Nagel J., Pukkala T., Skovsgaard J.P., Sotke R., Sterba, H., Recommendations for standardized documentation and further development of forest growth simulators, *Forstw. Cbl.* 121 (2002) 138–151.
- [42] Rawlings J.O., Sastry G.P., Dickey D.A., Applied Regression Analysis: A research tool, Springer-Verlag, New York, 1998
- [43] Reed D., Green E., Compatible stem taper and volume ratio equations, *For. Sci.* 30 (1984) 977–990.
- [44] Riemer T., Gadow K.v., Sloboda B., Ein Modell zur Beschreibung von Baumschäften, *Allg. Forst-Jagdztg.* 166 (1995) 144–147.
- [45] Rose C.E., Lynch T.B., Estimating parameters for tree basal area growth with a system of equations and seemingly unrelated regressions, *For. Ecol. Manage.* 148 (2001) 51–61.
- [46] Ryan T.P., Modern regression methods, John Wiley & Sons, New York, 1997.
- [47] SAS Institute Inc., SAS/STAT® 9.1.2, User's Guide, Cary, NC: SAS Institute Inc, 2004.
- [48] SAS Institute Inc., SAS/ETS® 9.1.2. User's Guide, Cary, NC: SAS Institute Inc., 2004.
- [49] SAS Institute Inc., SAS OnlineDoc® 9.1.2, Cary, NC: SAS Institute Inc., 2004.
- [50] Spurr S.H., Forest Inventory, The Ronald Press Co., New York, 1952.
- [51] Tarp-Johansen M.J., Skovsgaard J.P., Madsen S.F., Johannsen V.K., Skovgaard I., Compatible stem taper and stem volume functions for oak (*Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.) in Denmark, *Ann. Sci. For.* 54 (1997) 577–595.
- [52] Tassisa G., Burkhart H.E., Amateis R.L., Volume and taper equations for thinned and unthinned loblolly pine trees in Cutover, site-prepared plantations, *South. J. Appl. For.* 21 (1997) 146–152.
- [53] Teshome T., A ratio method for predicting stem merchantable volume and associated taper equations for *Cupressus lusitanica*, Ethiopia, *For. Ecol. Manage.* 204 (2005) 171–179.
- [54] Tomé M., Ribeiro F., Soares P., O modelo Globulus 2.1, Universidade Técnica de Lisboa-ISA, GIMREF, 2001.
- [55] Trincado G., Gadow K.v., Zur Sortimentschätzung stehender Laubbäume, *Cent.bl. Gesamte Forstwes.* 113 (1996) 27–38.
- [56] Trincado G., Gadow K.v., Tewari V.P., Comparison of three stem profile equations for *Quercus robur* L. *South Afr. For. J.* 177 (1996) 23–29.
- [57] Valdez J.R., Lynch T.B., Ecuaciones para estimar volumen comercial y total en rodales aclareados de pino patula en Puebla, México, *Agrociencia* 34 (2000) 747–758.
- [58] Van Deusen P.C., Sullivan A.D., Matney T.G., A prediction system for cubic foot volume of loblolly pine applicable through much of its range, *South. J. Appl. For.* 5 (1981) 186–189.
- [59] Van Deusen P.C., Matney, T.G., Sullivan A.D., A compatible system for predicting the volume and diameter of sweetgum tree to any height, *South. J. Appl. For.* 6 (1982) 159–163.
- [60] Vanclay J.K., Skovsgaard J.P., Evaluating forest growth models, *Ecol. Modell.* 98 (1997) 1–12.
- [61] West P.W., Ratkowsky D.A., Davis A.W., Problems of hypothesis testing of regressions with multiple measurements from individual sampling units, *For. Ecol. Manage.* 7 (1984) 207–224.
- [62] Xunta de Galicia, O monte Galego en cifras, Dirección Xeral de Montes e Medio Ambiente Natural, Santiago de Compostela, 2001.
- [63] Zellner A., An efficient method of estimating seemingly unrelated regressions and test for aggregation bias, *J. Am. Stat. Assoc.* 57 (1962) 348–368.
- [64] Zimmerman D.L., Núñez-Antón V., Parametric modelling of growth curve data: An overview (with discussion), *Test* 10 (2001) 1–73.