Eur J Forest Res (2005) 124: 177–186
DOI 10.1007/s10342-005-0066-6

ORIGINAL PAPER

A. Rojo · X. Perales · F. Sánchez-Rodríguez
J. G. Álvarez-González · K. von Gadow

Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain)

Received: 17 September 2004 / Accepted: 7 April 2005 / Published online: 1 September 2005

Abstract A total of 31 taper functions from 3 different groups of models (single, segmented and variable-form taper functions) were fitted to diameter-height data from 203 *Pinus pinaster* trees sampled across even-aged stands in Galicia (northwestern Spain). Most of the taper functions analyzed showed problems of multicollinearity as indicated by the condition number. A second-order autoregressive CAR(2) error process was incorporated into the models to minimize the effect of autocorrelation inherent in the longitudinal data used, and to provide valid tests of significance for model parameter estimates. In general, variable-form taper functions provided the most accurate predictions. The flexibility and predictive performance of the variable-form model developed by Kozak (For Chron 80(4):507–515, 2004) indicated its usefulness for estimating diameter at a specific height, merchantable volume, and total volume of Maritime pine in the study area.

Keywords Taper functions · Autocorrelation · Multicollinearity

Communicated by Hans Pretzsch

A. Rojo (✉) · X. Perales · J. G. Álvarez-González
Department of Ingeniería Agroforestal,
Escuela Politécnica Superior,
Universidad de Santiago de Compostela,
Campus universitario s/n, 27002 Lugo, Spain
E-mail: rojo@lugo.usc.es
Tel.: +34-982-252303
Fax: +34-982-285926

F. Sánchez-Rodríguez
Department of Producción Vegetal, Escuela Politécnica Superior,
Universidad de Santiago de Compostela,
Campus universitario s/n, 27002 Lugo, Spain

K. von Gadow
Institut für Waldinventur und Waldwachstum,
Georg-August-Universität Göttingen, Busgenweg 5,
37077 Gottingen, Germany

Introduction

Maritime pine (*Pinus pinaster* Ait.) is the most common forest species in Galicia (northwestern Spain), and occurs in both naturally regenerated stands and plantations. According to the third Spanish National Forest Inventory, almost 390,000 ha of land in Galicia are occupied by pure stands stocked with Maritime pine, and some 240,000 ha are occupied by mixed stands, mainly of eucalyptus and other pine or broad-leaved species; together these correspond to nearly 45% of the forested area in Galicia (Xunta de Galicia 2001).

Comparison of the data from the two most recent Spanish National Forest Inventories shows that *P. pinaster* cover increased slightly (by less than 5%) in Galicia between 1987 and 1998. The area occupied by mixed stands of *P. pinaster*, *Pinus radiata* and *Pinus sylvestris* has increased by almost 57%, whereas the area occupied by mixtures of *P. pinaster* and *Eucalyptus* spp. (mainly *E. globulus*) has increased by almost 17%.

In 1999 the total harvest of *P. pinaster* in Galicia was almost 2,300,000 m^3, which represented approximately 39% of the total regional harvest and almost 20% of the national Spanish timber harvest (Ministerio de Agricultura 1999). Wood of Maritime pine is mainly used for the chipboard and timber industries.

The Forest Plan of Galicia (Xunta de Galicia 1992) outlines *P. pinaster* as one of the species recommended for reforestation in the immediate future in the region, and furthermore that the Atlantic coast of Galicia should be considered for reforestation, mainly with this species of pine (Rodríguez-Soalleiro 1997).

There are some tools currently available for calculating the wood volume of Maritime pine stands, such as the equations used in the second National Forest Inventory, which are generally not of proven accuracy, and the volume equations included in some research studies (e.g. Rodríguez-Soalleiro 1995). However, none of these tools allows the estimation of merchantable volume at any stem diameter along the trunk in accordance with wood use in the industry, in spite of the importance of the species in Galicia.

This problem can be solved by developing stem taper functions or stem form equations (Kozak 1988; Newnham 1992; Riemer et al. 1995; Bi 2000). A stem form equation describes a mathematical relation between tree height and the stem diameter at that height. It is thus possible to calculate the stem diameter at any arbitrary height and conversely, to calculate the tree height for any arbitrary stem diameter. Consequently, the stem volume can be calculated for any log specification and it is possible to develop a volume equation for classified product dimensions (Castedo and Álvarez-González 2000; Gadow et al. 2001).

There are very few references to taper functions for *P. pinaster* in the available literature; one example is the study by Prieto and Tolosana (1991) and another that of Palma (1998), who used the taper equations developed by Kozak et al. (1969) and Demaerschalk (1972) to establish the influence of age on the stem form of the species sampled in the Portuguese coast. A taper function has not yet been developed for *P. pinaster* in Galicia.

The objective of the present study was to evaluate the performance of some well-known taper functions for *P. pinaster* in Galicia.

Materials and methods

Data

A total of 203 trees sampled from even-aged Maritime pine stands were used in the present study. The trees were taken from stands located throughout the area of distribution of Maritime pine in Galicia (Fig. 1), and included the existing range of diameters and heights (Fig. 2). Diameter at breast height (1.3 m above ground level) was measured to the nearest 0.1 cm in each tree. The trees were then felled leaving stumps of average height 0.11 m, and total bole length was measured to the nearest 0.01 m. Measurement intervals above breast height varied between 1 m and 2.5 m depending on the total tree height. In each section, two perpendicular diameters outside-bark were measured to the nearest 0.1 cm, and were then arithmetically averaged. Log volumes in cubic meters were calculated using Smalian's formula (Smalian 1837).

Figure 3 shows a plot of relative height against relative diameter, which are defined as the ratio between height above ground level to an upper-stem diameter and total tree height, and the ratio between an upper-stem diameter and diameter at breast height, respectively. A local regression curve with a smoothing parameter of 0.25 was fitted using the LOESS procedure of SAS/STAT (SAS Institute Inc. 2000a). This approach, pioneered by Cleveland et al. (1988), is flexible because no assumptions about the parametric form of the regression model are needed. The residuals of the nonparametric curve were examined for detecting abnormal data points (Bi 2000).

Models used

Many forms and types of stem profile models have been reported and evaluated for accuracy and precision (Sterba 1980). All of these models can be classified into three major groups: (1) single taper models; (2) segmented taper models and (3) variable form taper models.

Single taper models describe the entire profile of the stem using a single equation. The first models developed used lower order polynomials, in terms of the relative height on the stem. However, they were inadequate for describing the area near the base of the stem, therefore, high order polynomials were used to correctly characterize the butt swelling. Other single taper models use trigonometric functions to describe the bole taper. There are two characteristics that indicate that trigonometric functions may provide accurate stem profile models: (1) the analogy between trigonometric functions on the unit circle and the relative-height-relative diameter plots presented in many taper equations and (2) trigonometric functions can be expressed as Taylor's series of high-order polynomials (Thomas and Parresol 1991). Some single taper models use a power function of the relative height to define the stem profile. This approach was first introduced by Demaerschalk (1972) to develop a compatible volume system that ensures compatibility between the taper function and a total volume equation.

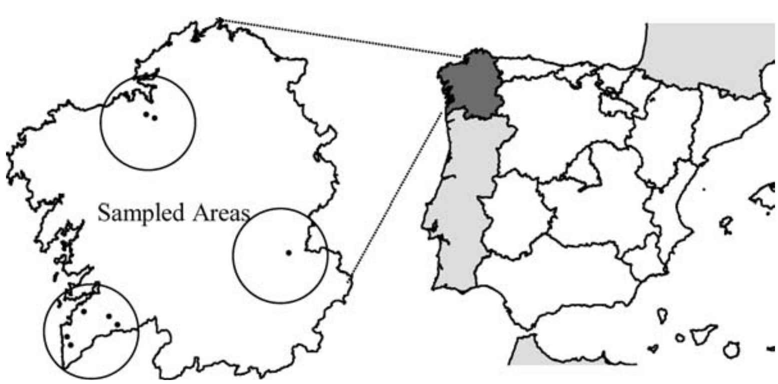

Fig. 1 Map showing the location of the sampled areas in Galicia and the location of Galicia within Spain

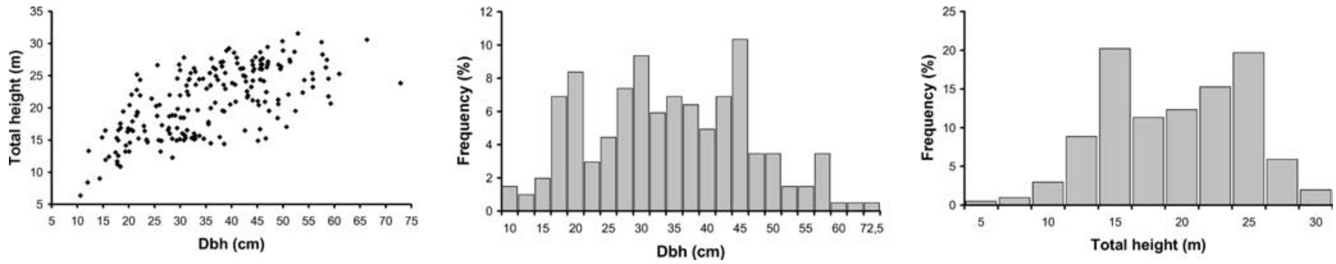

Fig. 2 Scatter plot and frequency of height and diameter of the 203 sample trees

A total of 20 single taper models were fitted in the present study; 10 of these are polynomials (Munro 1966; Bruce et al. 1968; two models proposed by Kozak et al. 1969; Bennett and Swindel 1972; Cervera 1973; Goulding and Murray 1976; Coffre 1982; Real and Moore 1986; and Jiménez et al. 1994), eight are power functions (Demaerschalk 1972; two models by Demaerschalk 1973; Ormerod 1973; two models by Newberry and Burkhart 1986; Reed and Green 1984; and Forslund 1990), one is a trigonometric function-based model (Thomas and Parresol 1991) and the last is an exponential model proposed by Biging (1984).

Relatively simple taper functions effectively describe the general taper of trees. However, they fail to characterise the entire stem profile and introduce bias, especially for the area near the butt and at the very top sections of tree (Jiang 2004). Max and Burkhart (1976) introduced the so-called segmented taper models to overcome the bias-induced poor performance of single taper models. These models use different sub-functions for various parts of the stem, conditioned to join smoothly, i.e. requiring the taper function to be C^2. The models selected were those of Max and Burkhart (1976), Valenti and Cao (1986), Parresol et al. (1987) and Farrar (1987). The first two models join three quadratic sub-functions at two join points, the model of Parresol et al. (1987) joins two cubic polynomials and the latter joins a power function with a combined-variable polynomial.

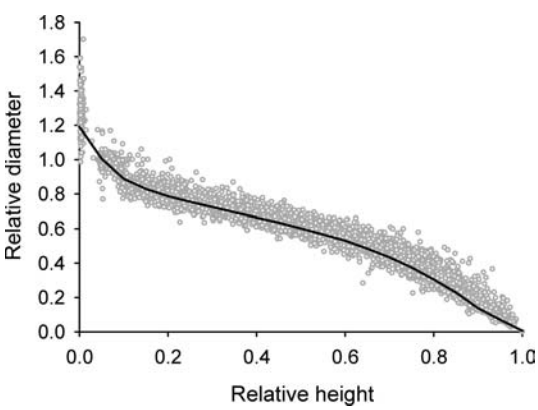

Fig. 3 Relative height plotted against relative diameter with a nonparametric local regression smoothing curve

A variable-form taper model describes the bole shape with a changing exponent or variable from ground to top to represent the neiloid, paraboloid, conic and several intermediate forms (Kozak 1997). This approach is based on the assumption that the stem form varies continuously along the length of a tree (Lee et al. 2003). In comparison with single and segmented taper models, this approach provides the lowest degree of local bias and greatest precision in taper predictions (e.g. Newnham 1988; Kozak 1988; Pérez et al. 1990; Muhairwe 1999). Seven variable-form taper models were analyzed in the present study: the model proposed by Riemer et al. (1995), two models proposed by Kozak (1988, 2004), two models proposed by Muhairwe (1999), the model proposed by Bi (2000), and the variable-exponent model proposed by Lee et al. (2003).

The mathematical expressions corresponding to each of the 31 taper functions that were used are shown in Table 1. The following notation will be used hereafter:

D Diameter at breast height outside bark (cm)

H Total height (m)

h Height above ground level (m)

d Diameter outside bark at a height h (cm)

b_i Coefficient to be estimated

X $((H - h)/(H - 1.30))$

T h/H

V_m Tree roundwood volume with bark (m^3). This variable has been obtained for each tree with the Smalian formula.

K $(\pi/40{,}000)$ is a constant to transform squared diameters into sections.

Z $((H-h)/(H))$

p (hi/H) where hi is the stem height of inflection point where the taper curve changes from neilod to parabolid.

According to Pérez et al. (1990) this inflection point is assumed to occur at between 15% and 35% of the total height. However, in the present study the parameter p was estimated.

Table 1 Analyzed taper functions

Model	Expression
Munro (1966)	$\left(\frac{d}{D}\right)^2 = b_1 - b_2 \cdot \left(\frac{h}{H-1.3}\right)$
Bruce et al. (1968)	$\left(\frac{d}{D}\right)^2 = \left(\frac{b_1 \cdot X^{1.5}}{10}\right) + \left(\frac{b_2 \cdot (X^{1.5}-X^3) \cdot D}{10^2}\right) + \left(\frac{b_3 \cdot (X^{1.5}-X^3) \cdot H}{10^3}\right) + \left(\frac{b_4 \cdot (X^{1.5}-X^{32}) \cdot H \cdot D}{10^5}\right) + \left(\frac{b_5 \cdot (X^{1.5}-X^{32}) \cdot \sqrt{H}}{10^3}\right) + \left(\frac{b_6 \cdot (X^{1.5}-X^{40}) \cdot H^2}{10^6}\right)$
Kozak et al. (a) (1969)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot (T-1) + b_2 \cdot (T^2-1)$
Kozak et al. (b) (1969)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot (1 - 2 \cdot T + T^2)$
Bennett and Swindel (1972)	$\left(\frac{d}{D}\right) = b_1 \cdot X + b_2 \cdot \frac{(H-h)\cdot(h-1.3)}{D} + b_3 \cdot \frac{(H-h)\cdot(h-1.3)\cdot H}{D} + b_4 \cdot \frac{(H-h)\cdot(h-1.3)\cdot(H+h+1.3)}{D}$
Cervera (1973)	$\frac{d}{D} = b_1 + b_2 \cdot X + b_3 \cdot X^2 + b_4 \cdot X^3 + b_5 \cdot X^4$
Goulding and Murray (1976)	$\frac{d^2 \cdot K \cdot H}{V} - 2 \cdot Z = \left[b_1 \cdot (3 \cdot Z^2 - 2 \cdot Z) + b_2 \cdot (4 \cdot Z^3 - 2 \cdot Z) + b_3 \cdot (5 \cdot Z^4 - 2 \cdot Z) + b_4 \cdot (6 \cdot Z^5 - 2 \cdot Z)\right]$
Coffre (1982)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot X + b_2 \cdot X^2 + b_3 \cdot X^3$
Real and Moore (1986)	$\frac{d^2}{D^2} = X^2 + b_1 \cdot (X^3 - X^2) + b_2 \cdot (X^8 - X^2) + b_3 \cdot (X^{40} - X^2)$
Jiménez et al. (1994)	$\left(\frac{d}{D}\right)^2 = b_1 + b_2 \cdot T + b_3 \cdot T^2 + b_4 \cdot T^3 + b_5 \cdot T^4 + b_6 \cdot T^5$
Demaerschalk (1972)	$d = b_1 \cdot D^{b_2} \cdot (H-h)^{b_3} \cdot H^{b_4}$
Demaerschalk (a) (1973)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot \left[\frac{(H-h)^{b_2}}{b_3 \cdot H^{b_2+1} + b_4 \cdot H^{b_2}}\right]$
Demaerschalk (b) (1973)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot \left(\frac{1}{D^2 \cdot H}\right) \cdot \left(\frac{H-h}{H}\right)^{b_2} + b_3 \cdot \left(\frac{H-h}{H}\right)^{b_4}$
Ormerod (1973)	$\frac{d}{D} = \left(\frac{H-h}{H-1.3}\right)^{b_1}$
Newberry and Burkhart (a) (1986)	$d = b_1 \cdot D \cdot (H-h)^{b_2}$
Newberry and Burkhart (b) (1986)	$d = b_1 \cdot D \cdot \left(\frac{H-h}{H-1.3}\right)^{b_2}$
Reed and Green (1984)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot \left(1 - \frac{h}{H}\right)^{b_2}$
Forslund (1990)	$\frac{d}{D} = \left(1 - T^{b_1}\right)^{1/b_2}$
Thomas and Parresol (1991)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot (T-1) + b_2 \cdot \sin(b_4 \cdot \pi \cdot T) + b_3 \cdot \text{cotan}\left(\frac{\pi \cdot T}{2}\right)$
Biging (1984)	$d = D \cdot \left[b_1 + b_2 \cdot \log\left(1 - T^{1/3}\right)\right] \cdot \left[1 - e^{(-b_1/b_2)}\right]$
Max and Burkhart (1976)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot (T-1) + b_2 \cdot (T^2-1) + b_3 \cdot (b_5 - T)^2 \cdot I_1 + b_4 \cdot (b_6 - T)^2 \cdot I_2 \quad I_1 = \begin{cases} 1 & \text{if } T \le b_5 \\ 0 & \text{if } T > b_5 \end{cases} \quad I_2 = \begin{cases} 1 & \text{if } T \le b_6 \\ 0 & \text{if } T > b_6 \end{cases}$
Valenti and Cao (1986)	$\left(\frac{d}{D}\right)^2 = b_1 \cdot Z + b_2 \cdot Z^2 + b_3 \cdot (Z - b_5) \cdot I_1 + b_4 \cdot (Z - b_6) \cdot I_2 \quad I_1 = \begin{cases} 1 & \text{if } Z \ge b_5 \\ 0 & \text{if } Z < b_5 \end{cases} \quad I_2 = \begin{cases} 1 & \text{if } Z \ge b_6 \\ 0 & \text{if } Z < b_6 \end{cases}$
Parresol et al. (1987)	$\left(\frac{d}{D}\right)^2 = Z^2 \cdot (b_1 + b_2 \cdot Z) + (Z - b_5)^2 \cdot [b_3 + b_4 \cdot (Z + 2 \cdot b_5)] \cdot I \quad I = \begin{cases} 1 & \text{if } Z \geqslant b_5 \\ 0 & \text{if } Z < b_5 \end{cases}$
Farrar (1987)	$d = D \cdot \left(\frac{h}{1.3}\right)^{b_5}$ if $h_t \le h \le 1.30$; $d = D \cdot X + b_1 \cdot \frac{(H-h)\cdot(h-1.3)}{H^2} + b_2 \cdot \frac{D \cdot (H-h)\cdot(h-1.3)}{H^2} + b_3 \cdot \frac{D^2 \cdot (H-h)\cdot(h-1.3)}{H^2} + b_4 \cdot \frac{(H-h)\cdot(h-1.3)\cdot(2 \cdot H - h - 1.3)}{H^3}$ if $1.30 \le h \le H$

Table 1 (Contd.)

Model	Expression
Riemer et al. (1995)	$d = \frac{b_1 \cdot D}{1-e^{b_3 \cdot (1.3-H)}} + \left(\frac{D}{2} - b_1 \cdot D\right) \cdot \left[1 - \frac{1}{1-e^{b_2 \cdot (1.3-H)}}\right] + e^{-b_2 \cdot h} \cdot \left[\frac{\left(\frac{D}{2}-b_1 \cdot D\right) \cdot e^{1.3 \cdot b_2}}{1-e^{b_2 \cdot (1.3-H)}}\right] - e^{b_3 \cdot h} \cdot \left[\frac{b_1 \cdot D \cdot e^{-b_3 \cdot H}}{1-e^{b_3 \cdot (1.3-H)}}\right]$
Kozak (1988)	$d = b_1 \cdot D^{b_2} \cdot b_3^D \cdot \left[\frac{1-\sqrt{T}}{1-\sqrt{p}}\right]^{\left[b_4 \cdot T^2 + b_5 \cdot \mathrm{Log}(T+0.001) + b_6 \cdot \sqrt{T} + b_7 \cdot e^T + b_8 \cdot \left(\frac{D}{H}\right)\right]}$
Muhairwe (a) (1999)	$d = b_1 \cdot D^{b_2} \cdot b_3^D \cdot \left[1 - \sqrt{T}\right]^{\left(b_4 \cdot T^2 + (b_5/T) + b_6 \cdot D + b_7 \cdot H + b_8 \cdot (D/H)\right)}$
Muhairwe (b) (1999)	$d = b_1 \cdot D^{b_2} \cdot \left[1 - \sqrt{T}\right]^{\left(b_3 \cdot T + b_4 \cdot T^2 + (b_5/T) + b_6 \cdot T^3 + b_7 \cdot D + b_8 \cdot (D/H)\right)}$
Bi (2000)	$\left(\frac{d}{D}\right) = \left[\log \sin\left(\frac{\pi}{2} \cdot T\right) / \log \sin\left(\frac{\pi}{2} \cdot \frac{1.3}{H}\right)\right]^{b_1 + b_2 \cdot \sin((\pi/2) \cdot T) + b_3 \cdot \cos((3 \cdot \pi/2) \cdot T) + b_4 \cdot \sin((\pi/2) \cdot T)/T + b_5 \cdot D + b_6 \cdot T \cdot \sqrt{D} + b_7 \cdot T \cdot \sqrt{H}}$
Lee et al. (2003)	$d = b_1 \cdot D^{b_2} \cdot (1 - T)^{b_3 \cdot T^2 + b_4 \cdot T + b_5}$
Kozak (2004)	$d = b_1 \cdot D^{b_2} \cdot H^{b_3} \cdot \left[\frac{1-T^{1/3}}{1-p^{1/3}}\right]^{\left[b_4 \cdot T^4 + b_5 \cdot \left(1/e^{D/H}\right) + b_6 \cdot \left(\frac{1-T^{1/3}}{1-p^{1/3}}\right)^{0.1} + b_7 \cdot \left(\frac{1}{D}\right) + b_8 \cdot H^{1-\left(\frac{h}{H}\right)^{1/3}} + b_9 \left(\frac{1-T^{1/3}}{1-p^{1/3}}\right)\right]}$

D Diameter at breast height outside bark (cm); *H* tree total height (m); *h* height above ground level (m); *d* diameter outside bark at a height *h* (cm); b_i, *p* coefficients to be estimated; V_m tree roundwood volume with bark (m^3); $X=((H-h)/(H-1.30))$; $T=h/H$; $K=\pi/40{,}000$ and $Z=((H-h)/(H))$

Multicollinearity and autocorrelation

Multicollinearity is defined as a high degree of correlation among several independent variables. This occurs when too many variables have been included in a model and a number of different variables measure similar phenomena. The existence of multicollinearity is not a violation of the assumptions underlying the use of regression, and therefore does not seriously affect the predictive ability of the model (Myers 1990; Kozak 1997). However, the presence of multicollinearity may inhibit the usefulness of the results as follows: (1) the variance of the predicted values tends to be inflated, especially when the values are not included in the sample used to fit the model, (2) the standard errors of the regression coefficients often have large variances with consequent lack of statistical significance, or have incorrect signs or are of the wrong magnitude (Myers 1990). The existence of multicollinearity is usual when developing taper functions with overcomplicated models that include several polynomial and cross-product terms.

To evaluate the presence of multicollinearity among variables in the models analyzed, the condition number was used; this is defined as the square root of the ratio of the largest to the smallest eigenvalue of the correlation matrix. The criteria for a condition number value that indicates serious multicollinearity are arbitrary, although a value of 30 is often quoted. Myers (1990) proposed a condition number of the correlation matrix higher than $\sqrt{1000}$ as indicative of serious multicollinearity.

In regression analysis, it is assumed that the error terms are independent, identically distributed, normal, random variables. However, construction of taper functions requires the collection of multiple observations for each tree (i.e., longitudinal data). Thus, it is reasonable to expect that the observations within each tree are spatially correlated, and that the assumption of independent error terms is violated. Although a statistical study of the correct analysis of the error structure of this kind of data exists (Zimmerman and Núñez-Antón 2001) it has often been ignored (Gregoire et al. 1995), mainly because the parameter estimates and the predicted values remain unbiased in the presence of autocorrelation (Kozak 1997). However, autocorrelation can have several adverse consequences in terms of the statistical inference when the aim is to identify statistically significant predictor variables (Garber and Maguire 2003).

Two general methods have been proposed to deal with continuous, unbalanced, multilevel longitudinal data. The first is to incorporate random subject effects (Gregoire et al. 1995), and the second is to model the correlation structure directly. In the present study a second-order continuous-time autoregressive error structure CAR(2) was used to overcome the inherent autocorrelation of the longitudinal data used. This error structure permits the model to be applied to irregularly spaced, unbalanced data (Gregoire et al. 1995; Zimmerman and Núñez-Antón 2001), both of which are characteristics of many forestry data sets (West et al. 1984). The CAR(2) model expands the error terms in the following way (Zimmerman and Núñez-Antón 2001):

$$e_{ij} = d_1 \rho_1^{h_{ij}-h_{ij-1}} e_{ij-1} + d_2 \rho_2^{h_{ij}-h_{ij-2}} e_{ij-2} + \varepsilon_{ij} \qquad (1)$$

where e_{ij} is the jth ordinary residual on the ith individual (i.e., the difference between the observed and the estimated diameters of the tree i at height measurement j), $d_k = 1$ for $j > k$ $k = 1,2$ and it is zero for $j \le k$, ρ_k are the autoregressive parameters to be estimated, and $h_{ij} - h_{ijk}$ are the distances separating the jth from the jth-k observations within each tree, $h_{ij} > h_{ij-k}$.

The main purpose of using the autocorrelation error structure is to obtain unbiased and efficient estimates of the parameters (Huang 1997; Parresol and Vissage 1998), and therefore the autocorrelation parameters ρ_k are generally ignored when using the model for predicting diameter and height. To evaluate the presence of autocorrelation and the effect of the CAR(2) error structure used, graphs representing residuals versus lag-residuals from previous observations within each tree were examined visually.

Fitting methodology and model validation

All the models were fitted by generalized least squares using the MODEL procedure of the SAS/ETS statistics programme (SAS Institute Inc. 2000b).

The accuracy and precision of diameter estimates of each model were compared using graphic and numeric analysis of the residuals (e_i). The plots of studentized residuals against the predicted diameter were examined for detection of possible systematic discrepancies. Three statistical criteria obtained from the residuals were also examined: bias ($\bar{E}$); mean square error (MSE) and the adjusted coefficient of determination (R^2_{adj}). These expressions may be summarized as follows:

$$\text{Bias} \quad \bar{E} = \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{n} \qquad (2)$$

$$\text{Mean square error} \quad MSE = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n - p} \qquad (3)$$

Adjusted coefficient of determination

$$R^2_{\text{adj}} = 1 - (n-1) \cdot \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n - p} \cdot \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \qquad (4)$$

where $e_i = y_i - \hat{y}_i$, and y_i,$\hat{y}_i$ and $\bar{y}_i$ are the measured, predicted and average values of the dependent variable, respectively; n is the total number of observations used to fit the model and p is the number of model parameters.

Akaike's information criterion differences ($AICd$), which is an index for selecting the best model that is based on minimizing the Kullback-Liebler distance, was used to compare models with a different number of parameters (Burnham and Anderson 1998):

$$AICd = n \cdot \ln \hat{\sigma}^2 + 2 \cdot (p+1) - \min(n \cdot \ln \hat{\sigma}^2 + 2 \cdot (p+1)) \qquad (5)$$

where p is the number of parameters of the model and $\hat{\sigma}^2$ is the estimator of the error variance of the model:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}.$$

Finally, a cross-validation approach was used to evaluate the prediction performance of the models. The bias, mean square error (MSE) and model efficiency of the estimates (ME), calculated by Eq. 4, were estimated using the residuals for fitting the model to a new data set obtained by deleting the observations of ten trees at a time selected from the same size category from the original data set. Although this approach is not a real method of model validation, it has been used as an additional criterion for selecting the best model and for detecting outliers (Myers 1990). Plots of the studentized residuals against the predicted diameter and plots showing the observed against the predicted diameter in cross-validation were also analyzed to detect systematic trends.

Results and discussion

After estimating the fit and cross-validation statistics of each one of the 31 models fitted, six models were selected for further analysis. In general, the variable-form models showed the best results, therefore, the four best variable-form taper equations (Kozak 1988; Riemer et al. 1995; Bi 2000 and Kozak 2004), the best simple taper function (Cervera 1973) and the best segmented model (Max and Burkhart 1976) were selected. The parameter estimates of each model are shown in Table 2, and the statistics of fit and cross-validation are given in Table 3.

The statistics in Table 3 and analyses of residual plots indicate that there are important differences among the three groups of models, but not among the variable-form taper models that showed the best performance in both fit and cross-validation phases. The taper function proposed by Kozak (2004) provided the best results on the basis of most statistics of fit and cross-validation.

The condition number clearly indicates the severity of multicollinearity in the models of Cervera (1973), Max and Burkhart (1976), and Kozak (1988), and to a lesser extent in Bi's (2000) model. However, the model proposed by Kozak (2004) and, especially the variable-form model proposed by Riemer et al. (1995) showed much lower multicollinearity. Similar results were found by Kozak (1997), who compared different variable-form taper equations.

Table 2 Values of the estimated parameters (ns means not significance at the 95% level)

Model	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	p
Cervera (1973)	0.0078	1.5014	1.1511	−5.3109	3.6929					
Max and Burkhart (1976)	−5.8940	2.9258	46.1267	−3.0050	0.1328	0.8511				
Riemer et al. (1995)	0.4006	0.6413	0.0985							
Kozak (1988)	0.8818	0.9947	0.9986	−0.7788	0.1250	−2.9989	1.6137	0.0692		0.1724
Bi (2000)	0.3537	0.3700	0.1185	−0.1089	−0.0004	0.0431	−0.1044			
Kozak (2004)	0.7914	0.9663	0.1095	0.4757	−0.4238	0.3894	1.2024	ns	0.2699	

Table 3 Values of the statistics in the fitting and validation step

Model	Parameters	Fitting					Validation			
		R^2_{adj}	*MSE*	$\bar{\bar{E}}$	Δ_j	Condition number	ME	MSE	$\bar{\bar{E}}$	Δ_j
Cervera (1973)	5	0.9852	2.8616	−0.0362	1016.03	380.33	0.9777	4.3028	−0.1175	1565.54
Max and Burkhart (1976)	6	0.9873	2.4527	0.0446	431.79	340.21	0.9804	3.7867	0.0877	1081.62
Riemer et al. (1995)	3	0.9883	2.2686	0.0902	132.73	5.33	0.9832	3.2398	0.1561	486.65
Kozak (1988)	9	0.9884	2.2331	−0.0442	78.90	358.53	0.9824	3.3932	−0.2604	668.18
Bi (2000)	7	0.9886	2.2093	−0.0562	36.29	168.75	0.9845	2.9799	−0.1762	245.70
Kozak (2004)	9	0.9887	2.1871	−0.0456	2982.9359	40.92	0.9853	2.8454	−0.4041	3981.3869

R^2_{adj} Adjusted coefficient of determination; *MSE* mean square error (cm^2); $\bar{\bar{E}}$ bias (cm); Δ Akaike's information criterion differences; *ME* model efficiency

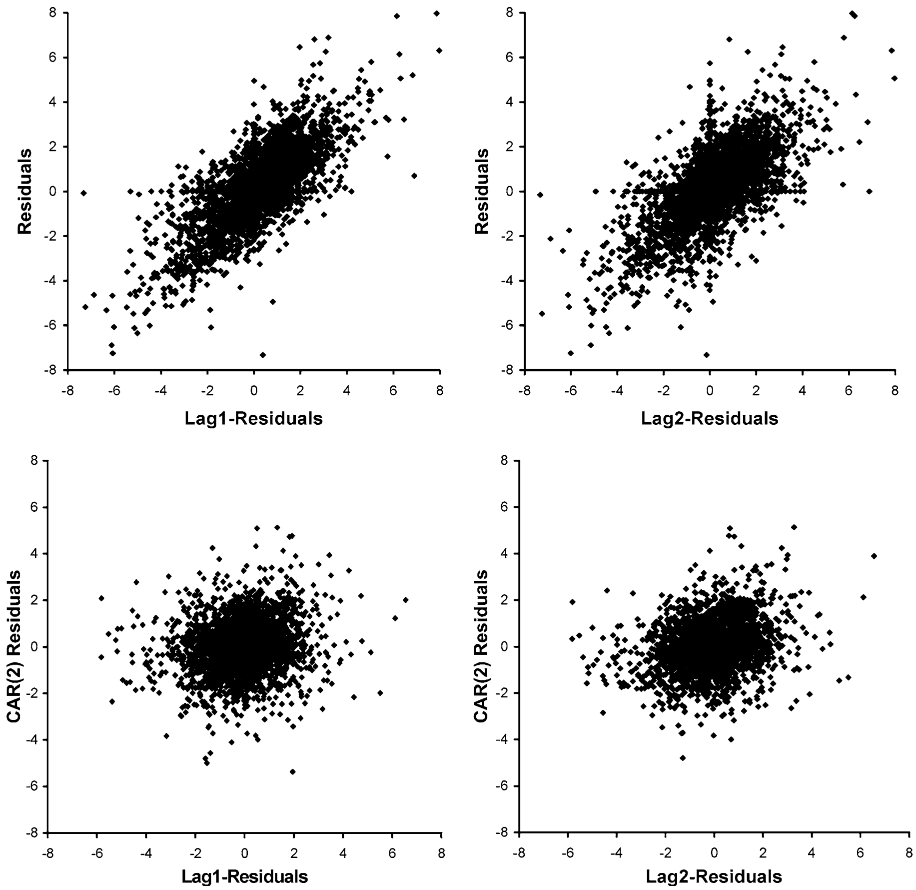

Fig. 4 Residuals as a function of: **a** Lag1-residuals (*left column*), **b** Lag2-residuals (*right column*) for both fitting methods: without error structure (*first row*), and assuming a CAR(2) error structure

Nonlinear fit of *P. pinaster* initially resulted in lag two autocorrelation in all the models, as expected because of the longitudinal nature of the data used for fitting. An example of the fit obtained with the Kozak (2004) model is shown in Fig. 4 (first row). Incorporation of the CAR(2) process removed the autocorrelation, as in-

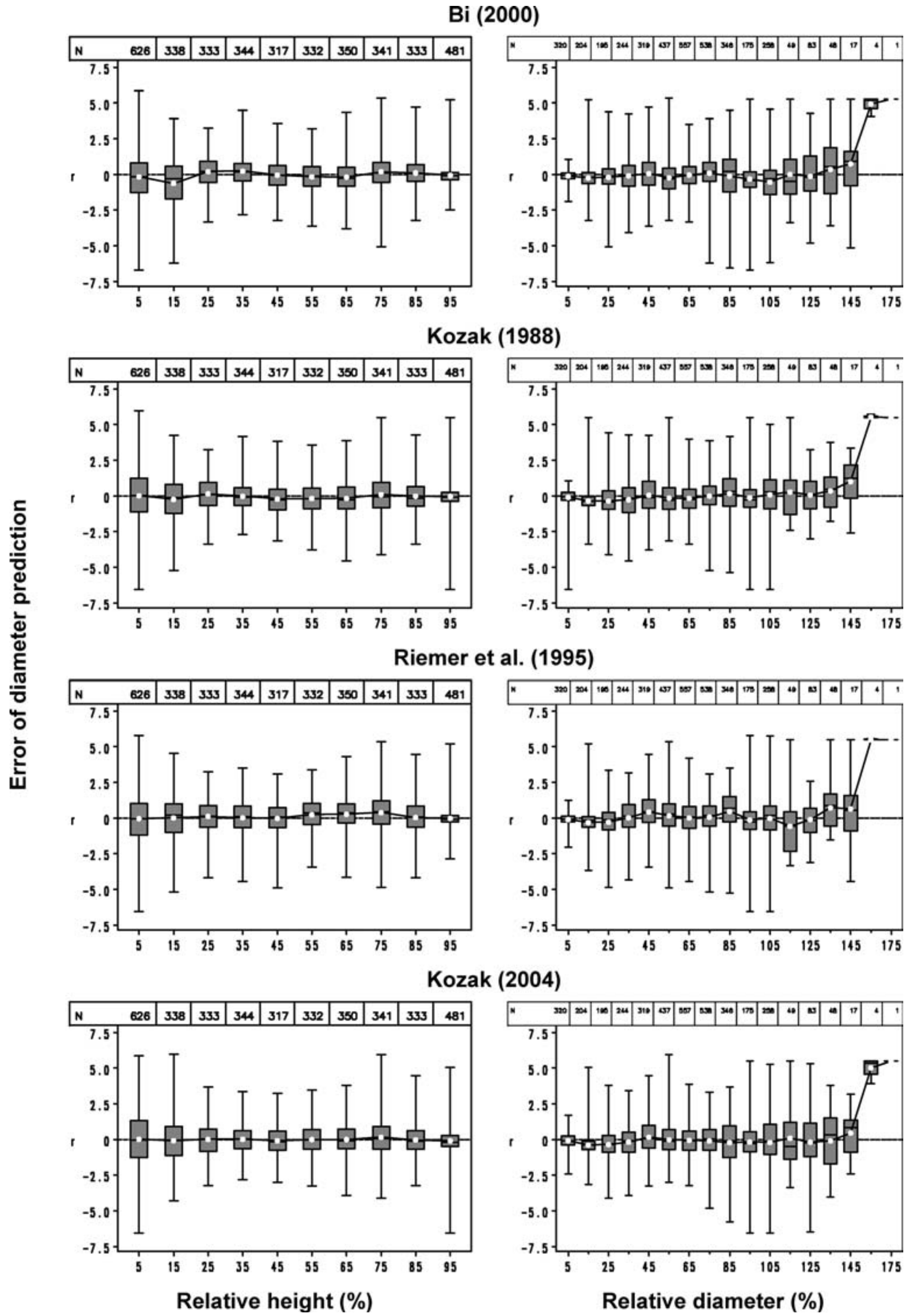

Fig. 5 Bias and precision of taper prediction for relative height and diameter, expressed as a percentage. The *white circles* represent the mean of diameter prediction error. The *box* represents the interquartile range. The maximum and minimum error are represented by the *upper and lower small horizontal segments* crossing the vertical bar, respectively. The *number above each bar* indicates the number of data points in the relative class

ferred from Fig. 4 (second row). The aim of autocorrelation correction was to improve the interpretation of the model's statistical properties, and has no use in practical applications.

Considering all of the factors analyzed, the most suitable taper models for developing a volume equation for Maritime pine in Galicia are the variable-form taper equations proposed by Kozak (1988), Riemer et al. (1995), Bi (2000) and Kozak (2004). The accuracy of diameter predictions by these four models was evaluated over relative height intervals and relative diameter intervals of 10%. The box-plots obtained for each taper function including the mean, maximum and minimum error of prediction, the median and the inter-quartile range are shown in Fig. 5.

There was little local bias across relative height classes in the diameter predictions obtained with the four models selected (Fig. 5, left row). The results from these

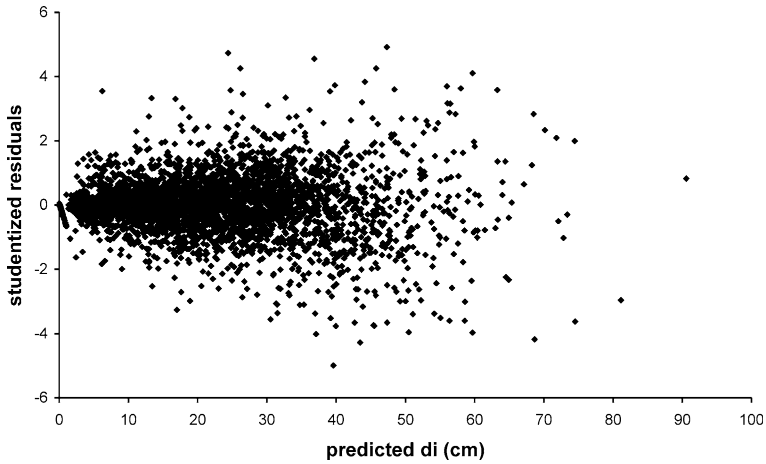

Fig. 6 Plot of studentized residuals against predicted diameter from the nonlinear least squares fits of variable-form taper function proposed by Kozak (2004)

four models were very similar, with an average size of error in diameter predictions below 0.3 cm for all relative height classes. The precision of prediction was relatively high as shown by the narrowness of the inter-quartile range across all relative height classes. As expected, the prediction corresponding to the section closest to the ground was generally less precise than that corresponding to other stem sections for the four models analyzed.

The box-plots of diameter predictions errors (Fig. 5, right row) showed a trend of increasing local bias across relative diameter classes. Again, the predictions corresponding to the lower stem (relative diameter upper 100%) were less precise, as indicated by the larger width of inter-quartile range for the four models. The bias was slightly negative for the upper stem (relative diameter classes ranged from 15 to 35%). The diameter predictions were particularly biased for the relative diameter classes up to 145%, with all the models showing serious overestimation of stem diameters because of the lack of data regarding old and bigger trees that are more neiloidal at the lower stem, i.e. with greater butt swell.

After the analysis of average values of the statistics obtained with the models and trends in bias along the stem for diameter and relative height, it was concluded that the most appropriate models for developing a volume equation with product classification for Maritime pine in Galicia are those of Kozak (1988), Riemer et al. (1995), Bi (2000) and Kozak (2004). All of these models provided similar results in terms of the bias trend along the stem. If the final choice of model is based on the fitting and validation statistics, especially Akaike's information criterion, the model of Kozak (2004) is suggested for use as taper function for *P. pinaster* in Galicia.

Because ordinary residuals are intrinsically not independent and do not have common variance, studentized residuals are used instead to overcome these problems (Rawlings 1988). The plot of studentized residuals against the predicted diameters obtained with Kozak's (2004) model is shown in Fig. 6. It is quite clear that the zero-studentized residuals cross the center of the data points and that the points do not show a trend of increasing error variances. This suggests that the taper function was appropriately identified and the error structure of the model is associated with the equal error variance (homoscedasticity).

Development of a volume equation

For the development of a volume equation it is necessary to use a function that describes the stem profile. The stem taper function should be integrable and, if it is possible, must have a generalized inverse $h = f(d)$.

Unfortunately, the model of Kozak (2004) is not analytically integrable and has no generalized inverse. Therefore, numerical integration methods (e.g. Kincaid and Cheney 1994) and iterative procedures to estimate height at a specific end diameter (e.g. Chapra and Canale 2002; Rade and Westergren 2004) should be used.

Acknowledgements The research reported in this paper was financially supported by the Spanish "Ministerio de Ciencia y Tecnología" through project AGL2001-3871-C02-01 ("Crecimiento y evolución de masas de pinar en Galicia"), and by the "Dirección Xeral de Investigación e Desenvolvemento" ("Xunta de Galicia") within the programme of grants for research stays in foreign institutions. We also thank two anonymous referees for their suggestions and useful comments and Dr. Christine Francis for revising the English grammar of the text.

References

Bennett FA, Swindel BF (1972) Taper curves for planted slash pine. USDA Forest Service Res. Note 179, p 4

Bi H (2000) Trigonometric variable-form taper equations for Australian Eucalypts. For Sci 46(3):397–409

Biging G (1984) Taper equations for second-growth mixed conifers of Northern California. For Sci 30(4):1103–1117

Bruce R, Curtiss L, Van Coevering C (1968) Development of a system of taper and volume tables for red alder. For Sci 14:339–350

Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. Springer, Berlin Heidelberg New York

Castedo F, Álvarez-González JG (2000) Construcción de una tarifa de cubicación con clasificación de productos para *Pinus radiata* D. Don en Galicia basada en una función de perfil del tronco. Invest Agrar: Sist Recur For 9(2):253–268

Cervera J (1973) El área basimétrica reducida, el volumen reducido y el perfil. Montes 174:415–418

Chapra SC, Canale RP (2002) Numerical methods for engineers. McGraw-Hill, Boston

Cleveland WS, Devlin SJ, Grosse E (1988) Regression by local fitting. J Econometrics 37:87–114

Coffre M (1982) Modelos fustales. Tesis Ing. For. Universidad Austral de Chile. p 44

Demaerschalk J (1972) Converting volume equations to compatible taper equations. For Sci 18(3):241–245

Demaerschalk J (1973) Integrated systems for the estimation of tree taper and volume. Can J For Res 3(1):90–94

Farrar RM (1987) Stem profile functions for predicting multiple product volumes in natural longleaf pines. South J Appl For 11(3):161–167

Forslund R (1990) The power function as a simple stem profile examination tool. Can J For Res 21:193–198

Gadow K von, Real P, Álvarez-González JG (2001) Modelización del crecimiento y la evolución de los bosques. IUFRO World Series Vol. 12, Vienna

Garber SM, Maguire DA (2003) Modelling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. For Ecol Manag 179:507–522

Goulding C, Murray J (1976) Polynomial taper equations that are compatible with tree volume equations. New Zealand J For Sci 5:313–322

Gregoire TG, Schabenberger O, Barrett JP (1995) Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Can J For Res 25:137–156

Huang S (1997) Development of a subregion-based compatible height-site index-age model for black spruce in Alberta. Alberta Land and Forest Service, For Manag Res Note no 5, Pub. no T/352. Edmonton, Alberta

Jiang L (2004) Compatible taper and volume equations for yellow-poplar in West Virginia. Master Thesis. West Virginia University Morgantown, West Virginia

Jiménez J, Aguirre O, Niembro R, Navar J, Domínguez A (1994) Determinación de la forma externa de *Pinus hartwegii* Lindl. en el noreste de México. Invest Agrar: Sist Recur For 3(2):175–182

Kincaid D, Cheney W (1994) Análisis numérico. Las matemáticas del cálculo científico. Addison-Wesley Iberoamericana, SA

Kozak A (1988) A variable-exponent taper equation. Can J For Res 18:1363–1368

Kozak A (1997) Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. Can J For Res 27:619–629

Kozak A (2004) My last words on taper equations. For Chron 80(4):507–515

Kozak A, Munro D, Smith J (1969) Taper functions and their application in forest inventory. For Chron 45(4):278–283

Lee WK, Seo JH, Son YM, Lee KH, Gadow K von (2003) Modeling stem profiles for *Pinus densiflora* in Korea. For Ecol Manag 172:69–77

Max T, Burkhart H (1976) Segmented polynomial regression applied to taper equations. For Sci 22:283–289

Ministerio de Agricultura (1999) Anuario de estadística agraria. Madrid

Muhairwe CK (1999) Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia. For Ecol Manag 113:251–269

Munro DD (1966) The distribution of log size and volume within trees. A preliminary investigation. University of British Columbia. Fac. of For., p 27

Myers RH (1990) Classical and modern regression with applications, 2nd edn. Duxbury Press, Belmont

Newberry J, Burkhart H (1986) Variable form stem profile models for Loblolly pine. Can J For Res 16:109–114

Newnham R (1988) A variable-form taper function. Fôrets Canada. Petawawa National Forestry Institute. Inf. Rep. PI-X-83, pp 1–33

Newnham R (1992) Variable-form taper functions for Alberta tree species. Can J For Res 22:210–223

Ormerod D (1973) A simple bole model. For Chron 49:136–138

Palma A (1998) Influência da idade na forma do perfil do tronco em Pinheiro Bravo (*Pinus pinaster* Aiton). Dunas do litoral portugués. Silva Lusitana 6(2):161–194

Parresol B, Vissage JS (1998) White pine site index for the southern forest survey. USDA For Serv Res Pap SRS-10

Parresol B, Hotvedt J, Cao Q (1987) A volume and taper prediction system for bald cypress. Can J For Res 17:250–259

Pérez D, Burkhart H, Stiff C (1990) A variable-form taper function for *Pinus oocarpa* Schiede. in Central Honduras. For Sci 36:186–191

Prieto A, Tolosana E (1991) Funciones de perfil para la cubicación de árboles en pie con clasificación de productos. Comunicaciones INIA. Serie Recursos Naturales no 58, Madrid

Rade L, Westergren B (2004) Mathematics handbook for science and engineering. Springer, Berlin Heidelberg New York

Rawlings JO (1988) Applied regression analysis. A Research Tool. Wadsworth, Belmont

Real PL, Moore JA (1986) An individual tree system for Douglas fir in the inland north-west. USDA Forestry Service General Technical Report NC-120, pp 1037–1044

Reed D, Green E (1984) Compatible stem taper and volume ratio equations. For Sci 30(4):977–990

Riemer T, Gadow K von, Slodoba B (1995) Ein Modell zur Beschreibung von Baumschäften. Allg Forst Jagdztg 166(7):144–147

Rodríguez-Soalleiro R (Coord.) (1997) Manual de Selvicultura del pino pinaster. Proyecto Columella. Área Forestal, Serie Manuales Técnicos. Escuela Politécnica Superior, Universidad de Santiago de Compostela

Rodríguez-Soalleiro R (1995) Crecimiento y producción de masas forestales regulares de *Pinus pinaster* Aiton. en Galicia. Alternativas selvícolas posibles. PhD Thesis. ETSI de Montes, Universidad Politécnica de Madrid. p 297

SAS Institute Inc (2000a) SAS/STAT User's guide version 8, Cary, NC

SAS Institute Inc (2000b) SAS/ETS User's guide version 8, Cary, NC

Smalian HL (1837) Beitrag zur Holzmeβkunst. Stralsund. In: Kramer H, Akça A (1995) Leitfaden zur Waldmeβlehre. JD Sauerländer's Verlag, Frankfurt am Main

Sterba H (1980) Stem curves—a review of the literature. Forest Abstr 41(4):141–145

Thomas C, Parresol B (1991) Simple, flexible, trigonometric taper equations. Can J For Res 21:1132–1137

Valenti M, Cao Q (1986) Use of crown ratio to improve loblolly pine taper equations. Can J For Res 16:1141–1145

West PW, Ratkowsky DA, Davis AW (1984) Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. For Ecol Manag 7:207–224

Xunta de Galicia (1992) Plan Forestal de Galicia. Síntese. Consellería de Agricultura, Gandería e Montes, Santiago de Compostela

Xunta de Galicia (2001) O monte galego en cifras. Dirección Xeral de Montes e Medio Ambiente Natural. Consellería de Medio Ambiente, Santiago de Compostela

Zimmerman DL, Núñez-Antón V (2001) Parametric modelling of growth curve data: an overview (with discussion). Test 10:1–73