Modelling diameter distributions of *Betula alba* L. stands in northwest Spain with the two-parameter Weibull function

J. J. Gorgoso[1]*, J. G. Álvarez González[2], A. Rojo[2] and J. A. Grandas-Arias[3]

[1] Departamento de Biología de Organismos y Sistemas. Universidad de Oviedo.
Escuela Universitaria de Ingenierías Técnicas. C/ Gonzalo Gutiérrez Quirós, s/n. 33600 Mieres. Spain
[2] Departamento de Enxeñaría Agroforestal. Universidade de Santiago de Compostela.
Escola Politécnica Superior. Campus Universitario. 27002 Lugo. Spain
[3] Dirección Xeral de Montes e Industrias Forestais. Consellería do Medio Rural. Xunta de Galicia.
San Lázaro, s/n. 15781 Santiago de Compostela. Spain

Abstract

The diameter distributions of 125 permanent plots installed in birch dominated (*Betula alba* L.) stands in Galicia were modelled with the two-parameter Weibull distribution. Four different fitting methods were used: that based on percentiles of the distribution, non linear regression, maximum likelihood and the method of moments. The most accurate fit was obtained with the non linear regression (NLR) approach, considering the following statistics in the comparison: bias, mean absolute error (MAE), mean square error (MSE) and number of plots rejected by the Kolmogoroff-Smirnoff (KS) test.

The scale parameter (*b*) and the shape parameter (*c*) obtained with the most accurate method (non linear regression), were first modelled with simple linear models and then related to commonly measured prediction variables (quadratic mean diameter, dominant height and stand density) with the parameter prediction model (PPM). The parameters fitted with the method of moments were recovered with the parameter recovery model (PRM) from the first and the second moments of the distribution (mean diameter and variance, respectively). Results indicated that both methods were successful in predicting the diameter frequency distributions. The PRM was more accurate than the PPM method.

Key words: diameter class model, two-parameter Weibull distribution, fitting methods, parameter modelling.

Resumen

Modelización de las distribuciones diamétricas en masas de *Betula alba* L. en el noroeste de España con la función Weibull biparamétrica

Las distribuciones diamétricas de 125 parcelas permanentes instaladas en masas puras de abedul (*Betula alba* L.) en Galicia fueron modelizadas con la distribución Weibull de dos parámetros. Se emplearon cuatro métodos de ajuste: basados en percentiles de la distribución, regresión no lineal, máxima verosimilitud y el método de los momentos. Los ajustes más precisos fueron obtenidos con regresión no lineal, considerando los siguientes estadísticos en la comparación de los resultados: sesgo, error medio absoluto, error medio cuadrático y número de parcelas rechazadas por el test de Kolmogoroff-Smirnoff.

El parámetro de escala (*b*) y el parámetro de forma (*c*) obtenidos con el método más preciso (regresión no lineal), fueron relacionados con variables de masa de frecuente medición (diámetro medio cuadrático, altura dominante y densidad) mediante modelos lineales sencillos aplicando la metodología de predicción de parámetros. Los parámetros ajustados con el método de los momentos fueron recuperados con modelos de recuperación de parámetros a partir del primer y del segundo momento de la distribución (diámetro medio y varianza, respectivamente). Los resultados indicaron que ambos métodos fueron satisfactorios para predecir las distribuciones de frecuencias de diámetros. El método de recuperación de parámetros fue más preciso que el método de predicción de parámetros.

Palabras clave: modelo de clases diamétricas, distribución Weibull biparamétrica, métodos de ajuste, modelización de parámetros.

Introduction

Birch (*Betula alba* L.) forests are a valuable natural resource in northwest Spain but a lack of studies about the profitability of the wood, as well as a lack of proper management have led to their general abandonment. At present birch stands cover 49,000 ha in northwest Spain (32,000 ha in Galicia and 17,000 ha in the adjoining region of Asturias) as dominant tree species, derived mainly from natural regeneration, frequently on abandoned land, and occasionally in plantations. Birch is the 5th species in terms of number of trees in Galicia, behind *Eucalyptus globulus*, *Pinus pinaster*, *Quercus robur* and *Quercus pyrenaica*. The total volume of Galician birch stands is nearly 3 million m³, which represents 2.3 % of the total wood produced in the region (Xunta de Galicia, 2001). These stands are of great ecological and protector value because they are usually located in hilly areas in the study region, under cold climates and adverse soil conditions.

Forest managers must respond to current demands and consider forests as multi-purpose sites (in terms of carbon storage, landscape value, as recreational sites, etc.); for this they need tools for evaluating different management practices and their effects on stand structure. Development of a growth models for the species (Rojo *et al.*, 2005; Diéguez-Aranda *et al.*, 2006) enables promotion of the productive and protective aspects of birch stands in northwest Spain. Diameter class models allow prediction of stand growth and planning of various uses and provide data about stand structure. These models are used to estimate stand variables and their structure with a density or distribution function, which is fitted to diameter distributions at breast height or individual tree volume.

There is, for example, great interest in knowing the number of trees in a diameter class in a stand in which a particular silvicultural treatment has been applied, because the diameter sizes determine the industrial use of the wood and thus the price of the different products. Diameter distributions also give information about stand structure, age structure, stand stability, etc. and enables planning of silvicultural treatments. Furthermore, tree diameter is an important factor in harvesting because it determines the type of machines used and how they perform during felling and transportation of the wood.

The Weibull probability density function was first used for modelling diameter distributions of pure and even-aged stands (Bailey and Dell, 1973), and since then it has been used in many growth models based on

diameter distributions because of its flexibility and simplicity (Rennolls *et al.*, 1985; Maltamo *et al.*, 1995; Kangas and Maltamo, 2000; Zhang *et al.*, 2003; Liu *et al.*, 2004).

In Spain different growth models based on diameter distributions have been developed, often by applying the two-parameter or the three-parameter Weibull function (Castilla and Prieto, 1992; Río, 1999; Álvarez-González *et al.*, 2002; García-Güemes *et al.*, 2002; Nanos and Montero, 2002; Palahí *et al.*, 2006a; Palahí *et al.*, 2006b).

The two-parameter Weibull function has been reported to be the most simple and accurate for modelling diameter distributions in birch stands (*Betula alba* L.) in northwestern Spain (Gorgoso, 2003), in accordance with Maltamo *et al.* (1995) for *Pinus sylvestris* and *Picea abies* stands in Finland, Álvarez-González (1997) for *Pinus pinaster* stands, and Condés (1997) for different species in Spain. These authors discussed the advantage of the use of the two-parameter Weibull probability density function over the three parameter function.

In this models two types of methods have been developed for modelling the parameters with stand variables (Hyink and Moser, 1983): Parameter prediction models (PPM) and parameter recovery models (PRM). Parameter prediction models (Smalley and Bailey, 1974; Álvarez-González, 1997; Torres-Rojo *et al.*, 2000) relate the parameters of a distribution function and stand variables by use of simple linear models. Parameter recovery models (PRM) were developed to improve the estimates obtained with the PPM models, and relate stand variables to percentiles (Bailey *et al.*, 1982; Cao and Burkhart, 1984) or moments (Newby, 1980; Burk and Newberry, 1984; Lindsay *et al.*, 1996; Río, 1999) of the diameter distribution.

The main aim of the study was to compare the accuracy of the two types of methodologies (PPM and PRM) with the two-parameter Weibull function and to compare four methods of fit: non linear regression, maximum likelihood, moments and percentiles, by applying them to data corresponding to 125 diameter distributions of *Betula alba* stands in northwest Spain.

Materials and Methods

Data set and processing

The data used to develop the diameter distribution model were obtained from a total of 125 research plots in *Betula alba* L. stands throughout Galicia (northwest

Spain), which were measured in 1998 and 1999. Birch is considered as a shade intolerant species in this area, so the stands are mainly even-aged. The size of the plots ranged from 200 m^2 to 1,000 m^2, depending on the stand density, in order to achieve a minimum of 30 trees per plot. The plots were established in birch-dominant stands (more than 85% of birch standing basal area), and were subjectively established to cover a wide variety of combinations of age, density and site.

All trees in each plot were numbered. Two perpendicular diameters at breast height were measured with callipers to the nearest 0.1 cm, and the arithmetic average was calculated. The minimum diameter inventoried was 5 cm. A total of 10,386 diameter measurements were available for analysis in the study. Furthermore, the heights of a random sample of 30 trees in each plot were measured to the nearest 0.25 m with a Blume-Leiss, for calculating the mean height. Dominant height was calculated from the percentage of the 100 thickest trees per ha (Assmann, 1970), and age was estimated by counting growth rings in samples extracted from some stems.

The following stand variables were calculated from the inventory data: age, density, quadratic mean diameter, basal area and mean and dominant height. Summary statistics including mean, maximum and minimum values and standard deviation of the main stand variables are shown in Table 1.

Other distribution variables of diameters were also calculated: mean diameter and the following percentiles: 16.731%, 25%, 50%, 75%, 63% and 97.366%, which were calculated by five different methods with SAS/STAT™ (SAS Institute Inc., 2001):

$$\text{Method 1:} \quad P_t = (1-g) \cdot x_{(j)} + g \cdot x_{(j} \qquad [1]$$

$$\text{Method 2:} \quad P_t = \begin{cases} x_{(i)} & \text{if } g \neq 0.5 \\ x_{(j)} & \text{if } g = 0.5 \text{ with } j \text{ even} \\ x_{(j+1)} & \text{if } g = 0.5 \text{ with } j \text{ odd} \end{cases} \qquad [2]$$

$$\text{Method 3:} \quad P_t = \begin{cases} x_{(j)} & \text{if } g = 0 \\ x_{(j+1)} & \text{if } g > 0 \end{cases} \qquad [3]$$

$$\text{Method 4:} \quad P_t = (1-h) \cdot x_{(k)} + h \cdot x_{(k+} \qquad [4]$$

$$\text{Method 5:} \quad P_t = \begin{cases} \dfrac{x_{(j)} + x_{(j+1)}}{2} & \text{if } g = 0 \\ x_{(j+1)} & \text{if } g > 0 \end{cases} \qquad [5]$$

where n is the number of data points, t is the estimated percentile per one, j is the entire part of the product of $n \cdot t$, i is the entire part of the product of $n \cdot t$ plus 0.5, g is the decimal part of the product $n \cdot t$, $x(j)$ the diameter sitting in the j position when they are ordered from minimum to maximum values, $x(j + 1)$ is the diameter in the position $j + 1$ with the same order, k is the entire part of the product $(n + 1) \cdot t$ and h is the decimal part of the product $(n + 1) \cdot t$.

Fitting methods

The two-parameter Weibull distribution (with the situation parameter $a = 0$) is obtained by integrating the Weibull density function, and has the following expression for a random variable x:

$$F(x) = \int_0^x \left(\frac{c}{b}\right) \cdot \left(\frac{x}{b}\right)^{c-1} \cdot e^{-\left(\frac{x}{b}\right)^c} \cdot dx = 1 - e^{-\left(\frac{x}{b}\right)^c} \qquad [6]$$

where $F(x)$ is the cumulative relative frequency of trees with diameter equal to or smaller than random variable, b is the scale parameter and c is the shape parameter.

Four methods of estimating the two-parameter Weibull distribution were compared: that based on percentiles of the distribution, non linear regression (NLR), maximum likelihood and moments.

Table 1. Summary of main stand variables for the sample used for modelling (N = 125)

Variable	Mean	Maximum	Minimum	Standard deviation
Age (years)	30.1	60	12	9.4
Density (trees · ha^{-1})	1,793	6,000	390	1,134.3
Quadratic mean diameter (cm)	14.1	23.2	7.4	3.5
Basal area (m^2 · ha^{-1})	24.6	66.5	3.3	10.7
Mean height (m)	13.1	21.3	6.5	3.1
Dominant height (m)	15	23.8	7.2	3.6

Method of percentiles

The method of percentiles for the estimation of the two parameter Weibull function was computed by Dubey (1967), Shiver (1988), Newberry *et al.* (1993) and Condés (1997). The values of the parameters were obtained with the following expressions:

$$\ln \hat{b} = \frac{\ln p_r - \dfrac{\ln p_t \cdot \ln(-\ln(1-r))}{\ln(-\ln(1-t))}}{1 - \dfrac{\ln(-\ln(1-r))}{\ln(-\ln(1-t))}} \qquad [7]$$

$$\hat{c} = \frac{\ln\left[\dfrac{\ln(1-r)}{\ln(1-t)}\right]}{\ln\left(\dfrac{p_r}{p_t}\right)} \qquad [8]$$

Shiver (1988) proposed values of r = 0.97366 and t = 0.16731, which were used in the present study.

Non Linear Regression (NLR)

The Weibull distribution is a non linear model, with the following general expression (Seber and Wild, 1989) for a dependent variable (*y*) and the independent variable (*x*):

$$y_i = m \cdot (x_i, \vec{\alpha}) + \varepsilon_i \qquad [9]$$

where *m* is a non linear function that depends on the parameter vector ($\vec{\alpha}$), which must be estimated, and ε_i the errors with the same criteria as the linear model (normal distribution, homocedasticity and independence). Estimation of the parameter vector ($\vec{\alpha}$) was carried out by minimum squares with the PROC NLIN procedure in SAS/STAT™ (SAS Institute Inc., 2001). The scale and shape parameters of the Weibull distribution were obtained with the non linear regression method by Álvarez-González (1997) and García-Güemes *et al.* (2002). The procedure was initiated with the values of the parameters obtained by the percentile method and the convergence of the parameters was obtained with the Gauss-Newton algorithm (Hartley, 1961).

Maximum Likelihood

The maximum likelihood estimation method used by Condés (1997), Nanos and Montero (2002) or Eerikäinen

and Maltamo (2003) allows calculation of the two distribution parameters with the following equations:

$$\frac{\sum\limits_{i=1}^{n}(x_i)^c \ln(x_i)}{\sum\limits_{i=1}^{n}(x_i)^c} - \frac{1}{c} = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i) \qquad [10]$$

$$b = \left(\frac{1}{n}\sum\limits_{i=1}^{n}(x_i)^c\right)^{\frac{1}{c}} \qquad [11]$$

where *n* equals the number of sample observations in a Weibull distribution and x_i (cm) the diameter of each tree. To calculate the two parameters (*b* and *c*) the LIFEREG procedure in SAS/STAT™ (SAS Institute Inc., 2001) was used.

Method of Moments

The method of moments used by Shifley and Lentz (1985), Nanang (1998) and Río (1999) is based on the relationship between the two parameters of the Weibull function and the first and second moments of the diameter distribution (mean diameter and variance, respectively):

$$b = \frac{\bar{d}}{\Gamma\left[1+\dfrac{1}{c}\right]} \qquad [12]$$

$$\sigma^2 = \frac{\bar{d}^2}{\Gamma^2\left[1+\dfrac{1}{c}\right]} \cdot \left(\Gamma\left[1+\dfrac{2}{c}\right] - \Gamma^2\left[1+\dfrac{1}{c}\right]\right) \qquad [13]$$

where $\bar{d}$ is the mean diameter of the distribution, σ^2 the variance and $\Gamma(i)$ is the Gamma function for each point (x = *i*). Equation [13] was resolved by an iterative procedure with the bisection method (Gerald and Wheatley, 1989).

Method comparison

The fitting method consistency was evaluated by the bias, mean absolute error (MAE), and mean square error (MSE), with the following expressions:

$$Bias = \frac{\sum\limits_{i=1}^{N} Y_i - \hat{Y}_i}{N} \qquad [14]$$

$$MAE = \frac{\sum_{i=1}^{N} |Y_i - \hat{Y}_i|}{N} \qquad [15]$$

$$MSE = \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N} \qquad [16]$$

where Y_i is the observed value, $\hat{Y}_i$ is the theoretical value predicted by the model and N is the number of data points.

Mean square error (MSE), mean absolute error (MAE) and bias were calculated for each fit in mean relative frequency of trees per one for all diameter classes and plots.

Each fit was also tested with the Kolmogoroff-Smirnoff (KS) test. This test compares the cumulative estimated frequency with the observed frequency. The most striking difference between the two distributions was the *D* value of the KS test: $D = \max[s(x) - F(x)]$, where S(x) is the cumulative frequency distribution observed for the sample x_i (i = 1, 2, ..., n) and F(x) the probability of the theoretical cumulative frequency distribution. The significance level was established at 20% and diameter classes of 1 cm intervals were chosen.

Modelling distribution parameters with stand variables

Correlation analysis was carried out by estimating Pearson's correlation coefficients between parameters corresponding to each plot and fitted by non linear regression and different stand variables or diameter distribution variables that can be estimated from stand variables with simple regression models. The variables chosen were: mean diameter of the distribution ($\bar{d}$), quadratic mean diameter of the stand (d_g), median (P_{50}), minimum diameter (D_{MIN}), maximum diameter (D_{MAX}), 25% percentile (P_{25}) and 75% percentile (P_{75}) of the distribution, density (N), basal area (G), dominant height (H_0) and median height (H_m) of the stand, and the natural logarithm of the cocient of the minimum diameter, the 25% percentile, the mean diameter and the median by the quadratic mean diameter (LD_{MIN}, LP_{25}, L_d and LP_{50}, respectively), in accordance with Álvarez-González (1997).

Pairs of parameters and variables were related with simple linear models by use of parameter prediction models (PPM). These equations were fitted simultaneously in order to correct errors because some variables are independent in some models and then become dependent, because they can not be obtained directly from yield tables. The simultaneous fitting was carried out with the FIML (full information maximum likelihood) method, by the PROC MODEL procedure in SAS/ETS (SAS Institute Inc., 2001).

On the other hand, parameters fitted by the moment's method were recovered with parameter recovery models (PRM), because the goodness of fit was better than that obtained with the percentiles method. With this method the mean diameter and the variance must be known (first and second order moments of the distribution, respectively). The variance was estimated with the following expression:

$$\text{var} = d_g^2 - \bar{d}^2 \qquad [17]$$

The mean diameter ($\bar{d}$) was related to the quadratic mean diameter (d_g), disaggregated from yield tables for birch stands in northwest Spain (Rojo *et al.*, 2005), and to stand variables that can be obtained from yield tables, with models in which mean diameter estimated is always smaller than quadratic mean diameter (Frazier, 1981):

$$\bar{d} = d_g - e^{X\beta} \qquad [18]$$

where X is the vector of independent stand variables in a fixed instant and β is the vector of parameters to be estimated.

For selecting the model, the following statistics obtained from residues were tested: bias [Eq. 14], mean square error (MSE) obtained in this case with the equation [19] and adjusted determination coefficient (R^2_{adj}):

$$MSE = \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N - p} \qquad [19]$$

$$R^2_{adj} = 1 - \frac{(N-1)\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{(N-p)\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \qquad [20]$$

where Y_i, $\hat{Y}_i$ and $\bar{Y}$ are, respectively, the observed, predicted and mean values of the dependent variable; N is the number of the data points in the fit and p is the number of parameters to be estimated.

Results

Fitting the two-parameter Weibull distribution

The mean values of bias, mean absolute error (MAE) and mean square error (MSE) obtained in the fits by the percentile approach, and the number of plots rejected by the KS test are shown in Table 2. The mean values of bias, MAE and MSE were similar for all methods of calculating percentiles. The number of plots rejected by the KS test was similar for the most accurate methods (approximately 34%).

The mean values of bias, MAE and MSE, and the number of events rejected by the KS test in the fits with non linear regression, maximum likelihood and moments are shown in Table 3. The most accurate fits, in terms of bias, MSE and number of rejections by the KS test (only 1 plot), were obtained with the non linear regression (NLR) approach and the least accurate fits were obtained with the maximum likelihood approach.

The results obtained with the method of moments were similar to those obtained with the percentiles method, and only better than NLR in terms of MAE. The number of events rejected by the KS test (42 plots), was 34% of the total, similar to percentage rejected by the percentiles and maximum likelihood approaches.

The mean values of bias and MSE for each diameter class obtained with the four methods of fitting are shown in Figure 1 and 2, respectively (in the fit with percentiles calculation method 2 was used because it was the most accurate). High values of bias were observed in the smallest diameter classes with

Table 3. Mean values of bias, mean absolute error in number of trees per one and number of plots rejected by the KS test ($\alpha = 0.20$) for three fitting methods for two-parameter Weibull distribution

Fitting method	Bias	MAE	MSE	KS test
Non Linear Regression	−0.0060	0.0177	0.0006	1
Maximum Likelihood	0.0102	0.0241	0.0014	40
Moments	0.0087	0.0156	0.0009	42

MAE: mean absolute error. MSE: mean square error. KS: number of rejections by the Kolmogoroff-Smirnoff test.

the maximum likelihood, moments and percentile approaches, and then the values decreased suddenly. The method that provided greatest stabilization of the bias throughout the diameter range was non linear regression (Fig. 1).

The maximum likelihood method provided high values of MSE up to a diameter of 26.5 cm, and then stabilized (Fig. 2). The non linear regression approach provided the most consistent results in terms of MSE, with values often less than 0.001 for each diameter class, always tending to decrease with increasing size of diameter classes; for the largest diameter class the results were similar to those obtained by the maximum likelihood approach, but lower than those based on percentiles and moments of the distribution. Nevertheless, in this diameter range the number of data points in the fits was less, with only 2.34% of trees with diameter larger than 25 cm. Furthermore, these methods were less accurate for the smallest diameter classes, and thus
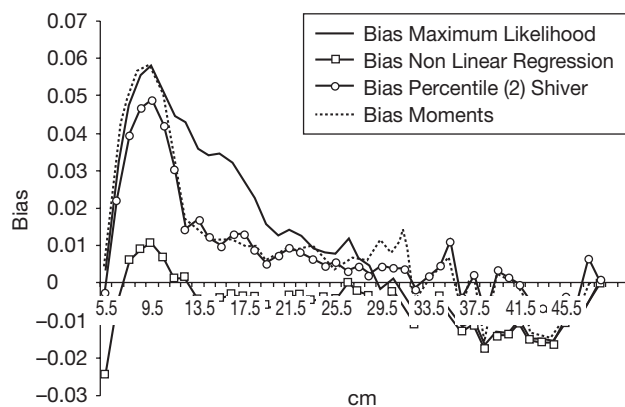
Table 2. Mean values of bias, mean absolute error, mean square error in number of trees per one and number of plots rejected by the KS test ($\alpha = 0.20$) for five different methods of calculation of percentiles

Method of calculation	Shiver ($P_{16.731}$, $P_{97.366}$)			
	Bias	MAE	MSE	KS test
Percentil (1)	0.0079	0.0150	0.0010	43
Percentil (2)	0.0085	0.0149	0.0010	44
Percentil (3)	0.0134	0.0172	0.0014	56
Percentil (4)	0.0166	0.0196	0.0015	60
Percentil (5)	0.0134	0.0172	0.0014	56

MAE: mean absolute error. MSE: mean square error. KS: number of rejections by the Kolmogoroff-Smirnoff test.

Figure 1. Mean value of bias in number of trees per one in each diameter class obtained by four fitting methods of the two-parameter Weibull distribution.
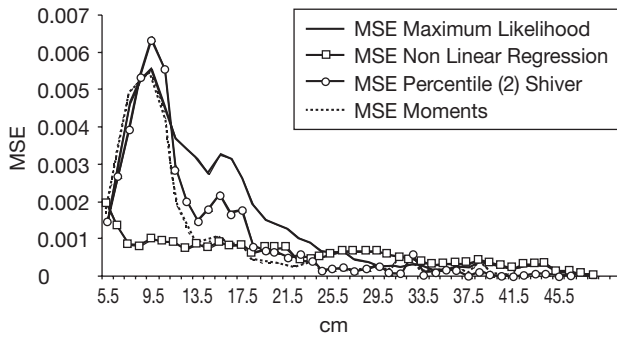
Figure 2. Mean value of mean square error (MSE) in number of trees per one in each diameter class obtained by four fitting methods of the two-parameter Weibull distribution.

the mean value of MSE for the total number of data was larger.

Modelling function parameters

The parameters fitted with the most accurate method (NLR) were predicted with the Parameter Prediction Model (PPM). Pearson's linear correlations for parameters b and c and stand variables and also diameter distribution variables are shown in Table 4.

The results show a good linear relationship between the scale parameter (b) and the measurement of central tendency. The value of the adjusted determination coefficient obtained was 0.99, in a model that relates b to the quadratic mean diameter (d_g).

$$b = -0.705 + 1.063 \cdot d_g \text{ (cm)} \quad R^2 = 0.99 \quad [21]$$

For the shape parameter (c) the highest value of Pearson's linear correlation (0.794) corresponded to the natural logarithm of the division between the 25% percentile and the quadratic mean diameter (LP_{25}).

Nevertheless the 25% percentile (P_{25}) cannot be obtained directly from yield tables and it was thus estimated by simultaneously fitting equation (22) with another that allows this percentile to be obtained from the density of the stand (N) and the dominant height

Table 4. Correlations between stand variables and parameters b and c of the Weibull function fitted with NLR

Variable	Parameter b	Parameter c
N	−0.597**	−0.362**
G	0.289**	−0.350**
H_0	0.633**	−0.232**
H	0.682**	−0.113
d_g	0.994**	0.228*
$\bar{d}$	0.970**	0.283**
D_{MAX}	0.706**	−0.206*
D_{MIN}	0.547**	0.722**
P_{25}	0.904**	0.561**
P_{50}	0.992**	0.316**
P_{75}	0.990**	0.135
LD_{MIN}	−0.412**	0.545**
LP_{25}	0.116	0.794**
L_d	0.020	0.260**
LP_{50}	0.332**	0.502**

** Significant at P < 0.01. * Significant at P < 0.05. *N:* density (trees · ha⁻¹). *G:* basal area (m² · ha⁻¹). *H₀:* dominant height (m). *H:* mean height (m). d_g: quadratic mean diameter (cm); $\bar{d}$: mean diameter (cm). D_{MAX}: Maximum diameter (cm). D_{MIN}: minimum diameter (cm). P_{25}: 25% percentile (cm). P_{50}: 50% percentile (cm). P_{75}: 75% percentile (cm); LD_{MIN}: ln(D_{MIN}/d_g). LP_{25}: ln(P_{25}/d_g). L_d: ln($\bar{d}/d_g$). LP_{50}: ln(P_{50}/d_g).

(H_0), with the FIML method (full information maximum likelihood).

$$c = 0.9597 - 0.7521 \cdot \cfrac{1}{\ln\left(\cfrac{P_{25}(cm)}{d_g(cm)}\right)} \quad R^2 = 0.89 \quad [22]$$

$$P_{25} = -1.8004 + 2.5726 \cdot \frac{100}{\sqrt{N}} + 0.3579 \cdot H_0 \quad R^2 = 0.63 \quad [23]$$

On the other hand, the parameter recovery model (PRM) was applied with the moments method of fitting. The results of the modelling of the mean diameter with the quadratic mean diameter (d_g) disaggregated from the yield tables for *Betula alba* in northwest Spain and with other stand variables are shown in Table 5.

Table 5. Parameter estimations and statistics used for modelling the mean diameter ($\bar{d}$) of the stand

Model	Par.	Estim.	Stand. error	Bias	MSE	R^2_{adj}
$\bar{d} = d_g - e^{b_0 + b_1 \cdot H_0 + b_2 \cdot t}$	b_0	−2.125	0.598	0.010	0.721	0.9409
	b_1	0.066	0.039			
	b_2	0.021	0.013			

Par.: parameter. Estim.: estimation. Stand. error: standard error. MSE: mean square error. R^2_{adj}: adjusted determination coefficient.

Table 6. Mean values of bias, mean absolute error, mean square error in number of trees per one and number of plots rejected by the KS test ($\alpha = 0.20$) for the two modelling methods: Parameter Prediction Model and Parameter Recovery Model

Fitting method	Modelling method	Bias	MAE	MSE	Rejections by the KS test
Non Linear Regression	Parameter Prediction Model (PPM)	–0.0030	0.0379	0.0041	36
Moments	Parameter Recovery Model (PRM)	0.0094	0.0332	0.0025	48

MAE: mean absolute error. MSE: mean square error. KS: Kolmogoroff-Smirnoff test.

The values of the statistics used in the comparison of the final results obtained with PPR fitted by non linear regression and PRM fitted by the method of moments after modelling the parameters are shown in Table 6. The mean values of bias and MSE in each diameter class obtained with PPR and PRM are shown in Figure 3.

Discussion

The large number of plots rejected by the KS test for the percentiles, maximum likelihood and moments methods implies that these methods are not appropriate for fitting diameter distributions of birch in northwest Spain. The results are consistent with those of García-Güemes *et al.* (2002), who obtained more accurate fits with the non linear regression method than with the method of percentiles for *Pinus pinea* stands in Valladolid (Spain).

Maximum likelihood, moments and percentile approaches underestimated the frequencies of trees in the smallest diameters class, although for the largest diameter class, the percentile approach was the most accurate. The maximum likelihood approach predicted larger values than the observed distribution for diameter classes of 30.5 cm and above, and never stabilized to values of bias close to zero, which is consistent with the poor results obtained by Condés (1997). Nevertheless, this method of fitting was applied successfully with the two-parameter Weibull model of Nanos and Montero (2002) in *Pinus pinaster* stands in Spain, and by Eerikäinen and Maltamo (2003) in *Pinus kesiya* stands in Zambia and Zimbabwe. Zhang *et al.* (2003) obtained better results with this method than with the moments and percentiles approaches in conjunction with the three-parameter function for mixed spruce-fir stands in northeastern North America. Palahí *et al.* (2006a, 2006b) fitted the left-truncated Weibull for different species in Catalonia for a truncation diameter of 7.5 cm. Parameters *b* and *c* were estimated by the maximum likelihood approach and the frequency of trees was predicted in trees/ha. High values of the root mean squared error were obtained for the smallest diameter classes.

In the case of the NLR approach, frequency distributions were often overestimated, with values of bias less than –0.02. The accuracy was lower than with the percentiles and method of moments for the largest diameter classes, but the shortage of data for these diameter classes implies less reliable results.
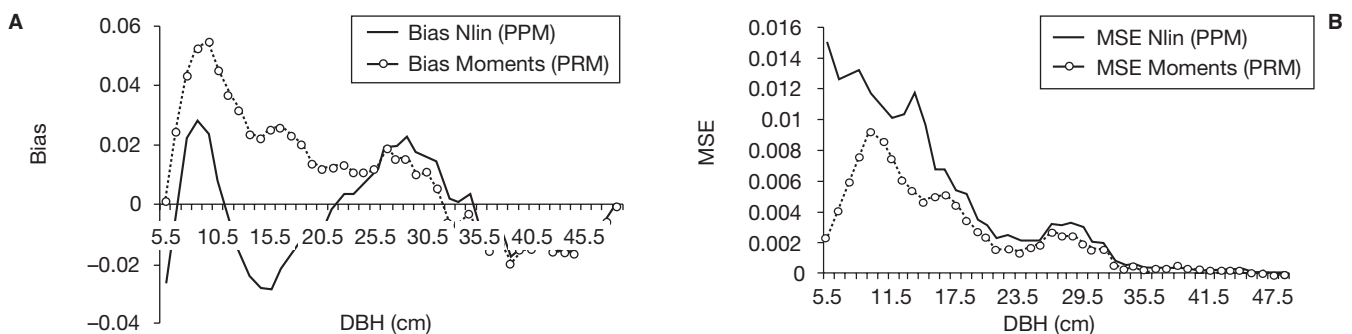
Figure 3 (A and B). Mean value of bias and MSE in number of trees per one in each diameter class with the PPM (Parameter Prediction Model) and PRM (Parameter Recovery Model).

The method of percentiles is very accurate for the largest diameter classes, which are of higher economic value. Nevertheless, because of the poor results for the smallest diameters and the large number of plots rejected by the KS test, this method is not recommended. The moments approach provided similar results, which were even better for some diameter classes. This is consistent with the findings of Nanang (1998), who obtained more accurate results with the method of moments than with the percentiles approach for *Azadirachta indica* plantations in Ghana. Bailey and Dell (1973) concluded that method of moments is more accurate than the percentile method, but pointed out that the calculations are more complicated.

Considering the changes in bias and MSE per diameter class (Figs. 1 and 2), the mean values of bias, MAE and MSE, and the number of rejections by the KS test (Tables 2 and 3), the non linear regression approach may be the most suitable for fitting the diameter distributions of birch stands in northwest Spain, with the two-parameter Weibull distribution.

As regards the parameter modelling, the parameter prediction model (PPM) avoids some of the inconveniences of the parameter recovery model (PRM) (Borders and Patterson, 1990; Nepal and Somers, 1992; Vanclay, 1995). Furthermore, the use of the two-parameter model avoids having to predict the situation parameter a ($a = 0$ in this case), which is always complex and not very accurate (Ortega, 1989; Maltamo *et al.*, 1995).

The Pearson's correlation analysis revealed close linear correlation between the scale parameter b and the quadratic mean diameter (d_g), in accordance with Álvarez-González (1997) and García-Güemes *et al.* (2002) who obtained linear models for this relationship, with values of adjusted determination coefficient (R^2_{adj}) close to 99%, higher than that obtained by Rennolls *et al.* (1985) for *Picea sitchensis* stands.

For modelling the shape parameter c by use of two simultaneous fits of linear models, the relationships between errors in the estimation and the endogenous error in the dependent variables have been considered (Torres-Rojo *et al.*, 2000). The independent variables in the models are quadratic mean diameter (d_g), density (N) and dominant height (H_0), which are all easily obtained from yield tables. In both models the values of adjusted determination coefficient (R^2_{adj}) were high (0.89 and 0.63, respectively) considering that the prediction of this parameter is often not very accurate. Rennolls *et al.* (1985) and Maltamo *et al.* (1995) for

example, only explained a small part of the total variability in the parameter (5% and a 31% respectively), in the first case with a parabolic model including the mean diameter, and in the second case, with a linear model including the quadratic mean diameter and the natural logarithm of the age. Palahí *et al.* (2006a) developed parameter prediction models for *Pinus sylvestris*, *Pinus nigra* and *Pinus halepensis* in Catalonia using the truncated Weibull function. These authors used regression analysis to establish the relationship between Weibull parameters and stand basal area, number of trees per hectare and elevation of the site.

When the parameters were recovered, the most accurate model for modelling mean diameter was that predicting the mean diameter without the stand density (N), which was not significant (at a level of 5%). Thus mean diameter was modelled with quadratic mean diameter (d_g), dominant height (H_0) and age (t). In this case the value of the adjusted determination coefficient (R^2_{adj}) was 0.94.

The final results showed that PRM provided more accurate estimates than PPM in terms of MSE and MAE (see Table 6), although the KS test rejected more of the events. Thus, the parameter modelling with PRM was more accurate because similar final values were obtained in terms of the number of events rejected by the KS test for both methods in relation to the large differences in the fits (one plot rejected by NLR approach and 42 plots with the method of moments). Recently Liu *et al.* (2004) found that PRM performed better than PPM in terms of predicting error for plantations of black spruce (*Picea mariana*) in central Canada.

In conclusion, the NLR approach was found to be more accurate than method of moments, maximum likelihood and percentiles for fitting the two-parameter Weibull distribution to diameter distributions of birch in northwest Spain. Parameter prediction models and parameter recovery models were successful methods of modelling the function parameters, with the PRM being more efficient for estimating the theoretical distribution.

Acknowledgements

Folgueira, Joaquín Pedre and Darío Ferreiro for help in analysing the data set.

References

ÁLVAREZ-GONZÁLEZ J.G., 1997. Análisis y caracterización de las distribuciones diamétricas de *Pinus pinaster* Ait. en Galicia. Ph D thesis, Univ Politécnica de Madrid (Unpublished).

ÁLVAREZ-GONZÁLEZ J.G., SCHRÖDER J., RODRÍGUEZ R.J., RUÍZ A.D., 2002. Modelling the effect of thinnings on the diameter distribution of even-aged Maritime pine stands. For Ecol Manage 165, 57-65

ASSMANN E., 1970. The Principles of Forest Yield Study. Pergamon Press, Oxford, New York.

BAILEY R.L., DELL T.R., 1973. Quantifying Diameter Distributions with the Weibull Function. For Sci 19, 97-104.

BAILEY R.L., ABERNETHY N.C., JONES E.P., 1982. Diameter distribution models for repeatedly thinned slash pine plantations. USDA For Ser Gen Tech Re SO-34.

BORDERS B.E., PATTERSON W.D., 1990. Projecting stand tables: a comparison of the Weibull diameter distribution method, a percentile-based projection method, and a basal area growth projection method. For Sci 36, 413-424.

BURK T.E., NEWBERRY J.D., 1984. A simple algorithm for moment-based recovery of Weibull distribution parameters. For Sci 30, 329-332.

CAO Q.V., BURKHART H.E., 1984. A segmented distribution approach for modeling diameter frequency data. For Sci 30, 129-137.

CASTILLA G., PRIETO A., 1992. Desarrollo y utilización de un simulador de producción de madera para masas de *Pinus radiata* en el País Vasco. Invest Agrar: Sist Recur For 1, 189-209.

CONDÉS S., 1997. Simulación de parcelas arboladas con datos del II Inventario Forestal Nacional. Ph D thesis, Univ. Politécnica de Madrid (Unpublished).

DIÉGUEZ-ARANDA U., GRANDAS-ARIAS J.A., ÁLVAREZ-GONZÁLEZ J.G., GADOW K.v., 2006. Site quality curves for birch stands in north-western Spain. Silva Fennica 40 (4), 631-644.

DUBEY S.D., 1967. Some percentile estimators for Weibull parameters. Technometrics 9, 119-129.

EERIKÄINEN K., MALTAMO M., 2003. A percentile based basal area diameter distribution model for predicting the stand development of *Pinus kesiya* plantations in Zambia and Zimbabwe. For Ecol Manage 172, 109-124.

FRAZIER J.R., 1981. Compatible whole-stand and diameter distribution models for loblolly pine. Ph D thesis, VPI & SU (Unpublished).

GARCÍA-GÜEMES C., CAÑADAS N., MONTERO G., 2002. Modelización de la distribución diamétrica de las masas de *Pinus pinea* L. de Valladolid (España) mediante la función de Weibull. Invest Agr: Sist Recur For 11, 264-282.

GERALD C.F., WHEATLEY P.O., 1989. Applied numerical analysis. (4th edition). Addison-Wesley publishing Co, Reading, Massachusetts.

GORGOSO J.J., 2003. Caracterización de las distribuciones diamétricas de *Betula alba* L. en Galicia. Ph D thesis, Univ Santiago de Compostela (Unpublished).

HARTLEY H.O., 1961. The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. Technometrics 3, 269-280.

HYINK D.M., MOSER J.W., 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. For Sci 29, 85-95.

KANGAS A., MALTAMO M., 2000. Performance of percentile based diameter distribution prediction and Weibull method in independent data sets. Silva Fennica 34, 381-398.

LINDSAY S.R., WOOD G.R., WOOLLONS R.C., 1996. Modelling the diameter distribution of forest stands using the Burr distribution. Journal of Applied Statistics 23, 609-619.

LIU C., ZHANG S.Y., LEI Y., NEWTON P.F., ZHANG L., 2004. Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. Can J For Res 34, 2424-2432.

MALTAMO M., PUUMALAINEN J., PÄIVINEN R., 1995. Comparison of beta and Weibull functions for modelling basal area diameter distribution in stands of *Pinus sylvestris* and *Picea abies*. Scan J For Res 10, 284-295.

NANANG D.M., 1998. Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of neem plantations in Northern Ghana. For Ecol Manage 103, 1-7.

NANOS N., MONTERO G., 2002. Spatial prediction of diameter distributions models. For Ecol Manage 161, 147-158.

NEPAL S.K., SOMERS G.L., 1992. A generalized approach to stand table projection. For Sci 38, 120-133.

NEWBY M., 1980. The properties of moment estimators for the Weibull distribution based on the sample coefficient of variation. Technometrics 22, 187-194.

ORTEGA A., 1989. Modelos de evolución de *Pinus Sylvestris* L. Ph D thesis. Univ Politécnica de Madrid (Unpublished).

PALAHÍ M., PUKKALA T., TRASOBARES A., 2006a. Modelling the diameter distribution of *Pinus sylvestris*, *Pinus nigra* and *Pinus halepensis* forest stands in Catalonia using the truncated Weibull function. Forestry 79(5), 553-562.

PALAHÍ M., PUKKALA T., TRASOBARES A., 2006b. Calibrating predicted tree diameter distributions in Catalonia, Spain. Silva Fennica 40(3), 487-500.

RENNOLLS K., GEARY D.N., ROLLINSON T.J., 1985. Characterizing diameter distributions by the use of the Weibull Distribution. Forestry 58, 57-66.

RÍO M. del, 1999. Régimen de claras y modelo de producción para *Pinus sylvestris* L. en los sistemas Central e Ibérico. Tesis doctorales INIA, Serie: Forestal 2. Madrid, España.

ROJO A., ÁLVAREZ-GONZÁLEZ J.G., GRANDAS J.A., DIÉGUEZ-ARANDA U., 2005. Tablas de producción de selvicultura media para el abedul (*Betula alba* L.) en Galicia. IV Congreso Forestal Español. Zaragoza, 26-30 septiembre 2005.

SAS INSTITUTE INC., 2001. SAS/ETS User's Guide, Version 8.2. Cary, North Carolina,

SAS INSTITUTE INC., 2001. SAS/STAT™ User's Guide, Version 8.2. Cary, North Carolina.

SEBER G.A.F., WILD C.J., 1989. Nonlinear regression. Wiley and Sons, New York, U.S.A..

SHIFLEY S.R., LENTZ E.L., 1985. Quick estimation of the three-parameter Weibull to describe tree size distributions. For Ecol Manage 13, 195-203.

SHIVER B.D., 1988. Sample size and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions. For Sci 34, 809-814.

SMALLEY G.W., BAILEY R.L., 1974. Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama and Georgia highlands. USDA. For Serv Res Pap SO-96.

TORRES-ROJO J.M., MAGAÑA O.S., ACOSTA M., 2000. Metodología para mejorar la predicción de parámetros de distribuciones diamétricas. Agrociencia 34, 627-637.

VANCLAY J.K., 1995. Growth models for tropical forests: A synthesis of models and methods. For Sci 41, 7-42.

XUNTA DE GALICIA, 2001. O monte galego en cifras. Consellería de Medio Ambiente, Dirección Xeral de Montes e Medio Ambiente Natural. 226 pp.

ZHANG L., PACKARD K.C., LIU C., 2003. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce-fir stands in northeastern North America. Can J For Res 33, 1340-1347.