



**HAL**  
open science

# Modelling the diameter distribution of eucalyptus plantations with Johnson's S B probability density function: parameters recovery from a compatible system of equations to predict stand variables

Mateus, Tomé

## ► To cite this version:

Mateus, Tomé. Modelling the diameter distribution of eucalyptus plantations with Johnson's S B probability density function: parameters recovery from a compatible system of equations to predict stand variables. *Annals of Forest Science*, 2011, 68 (2), pp.325-335. 10.1007/s13595-011-0037-7. hal-00930772

**HAL Id: hal-00930772**

**<https://hal.science/hal-00930772>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modelling the diameter distribution of eucalyptus plantations with Johnson's $S_B$ probability density function: parameters recovery from a compatible system of equations to predict stand variables

Ayana Mateus · Margarida Tomé

Received: 6 April 2010 / Accepted: 3 November 2010 / Published online: 2 March 2011  
© INRA and Springer Science+Business Media B.V. 2011

## Abstract

- **Introduction** The simulation of diameter distributions is the basis for predicting volume in the so-called diameter distribution models. Combined with volume, volume ratio and taper equations, these models allow the prediction of volume assortments according to user needs. The simulation of diameter distributions is also essential in initialising individual tree models. It is also a useful aid for planning harvesting operations.
- **Methods** In this paper, Johnson's distribution was used to model the diameter distribution of *Eucalyptus globulus* in Portugal. When a predefined probability density function is used as part of a growth and yield model, the parameters of the function must be estimated for each year during the simulation period.
- **Results** The development of a system of equations that relates stand characteristics to mathematical

functions of the distribution, such as the moments of the distribution, allows for the estimation of parameters (i.e. parameter recovery). This method assures compatibility between the characteristics of the observed population used in parameter recovery and those obtained through simulation.

- **Conclusions** The system of equations was built in such a way that the observed well-established biological processes between stand variables is maintained, and the equations were simultaneously fitted to minimise the determinant of the covariance matrix of errors. Based on validation with an independent data set, the model provides precise estimates of total stand volume.

**Keywords** Probability density function · Johnson's  $S_B$  distribution · Diameter distribution · Forest planning

## 1 Introduction

Research on the development of growth and yield models has included the use of probability density functions to model diameter distributions, such as the log-normal, gamma, beta, Johnson's  $S_B$  and Weibull functions (Aranda 2004; Bailey and Dell 1973; Hafley and Buford 1985; Fonseca et al. 2009; Hafley and Schreuder 1977; Kamziah et al. 1999; Kiviste et al. 2003; Li et al. 2002; Maltamo et al. 1995; Palahi et al. 2007; Parresol 2003; Rennolls and Wang 2005; Zhang et al. 2003; Zhou and McTague 1996).

More recently, (Palahi et al. 2007) compared the fit of the beta, Johnson's  $S_B$ , Weibull and truncated Weibull functions to diameter distributions of forest

---

Handling Editor: Reinhart Ceulemans

A. Mateus (✉)  
CMA, Departamento de Matemática,  
Faculdade de Ciências e Tecnologia, FCT,  
Universidade Nova de Lisboa,  
Quinta da Torre, 2829-516 Caparica, Portugal  
e-mail: amf@fct.unl.pt

M. Tomé  
Centro de Estudos Florestais (CEF),  
Instituto Superior de Agronomia (ISA),  
Universidade Técnica de Lisboa (UTL),  
Tapada da Ajuda, 1349-017 Lisbon, Portugal  
e-mail: magatome@isa.utl.pt

stands in Catalonia. These distribution functions were fit to the observed diameter distributions of the number of stems and the stand basal area based on truncated data from the National Forestry Inventory, with the diameter threshold equal to 7.5 cm. In all cases, the truncated Weibull function for the diameter distribution of the stand basal area appeared to be the most accurate and consistent function. The beta and Johnson's  $S_B$  were the second best and nearly equally good.

The simulation of diameter distributions is also essential to initialise individual tree models (Ek and Monserud 1979), including, for instance, the simulation of the transition from even-aged forestry, simulated with whole stand models, to multi-objective or close-to-nature forestry, which requires the use of individual tree models. The initialisation of individual tree models may also be needed when raw data are not available from forest inventory. Additionally, diameter distribution simulation is an essential aid to more efficiently plan harvesting operations, which usually represent a high percentage of the costs associated with pulp production. When a probability density function is used as part of a growth and yield model, there arises the need to estimate the parameters of the function for each year during the simulation period. In early work, the parameters were estimated directly as regression functions of stand characteristics (e.g. age, site quality and stand density). Due to the relationships among parameters and to some inconsistency in the relationships of the parameters with stand characteristics, it was apparent that improvements were needed. Therefore, (Hyink and Moser 1983) presented a new methodology according to which stand characteristics, such as the quadratic mean diameter and the basal area, are first estimated, and then the parameter estimates are obtained by relating these characteristics with mathematical functions based on the probability density function under consideration. The mathematical functions used by these authors included the distribution moments. These techniques have been referred to as parameter recovery methods. The main advantage of using parameter recovery methods is that the stand variables that are used in the parameter recovery assure compatibility between the characteristics of the stand and those obtained through simulation. Additionally, this method usually yields a better prediction of stand volume than methods previously used (Hafley and Schreuder 1977; Reynolds et al. 1988).

*Eucalyptus globulus* Labill. is one of the most important economic forest species in Portugal, occupying an area of 875,000 ha of a total forest area of 3,346,000 ha (Tomé et al. 2007). It is a fast-growing species that is mainly used commercially by the pulp industry. The

trees are planted at final density, as thinning and pruning practices are unusual in first-rotation stands. The stands are intensively managed in a short-rotation coppice system in which the first cycle is from single-stem planted seedlings and is followed by two or three coppiced stands with an average cutting cycle of 10–12 years.

The objective of the research reported here was to model the diameter distribution of first-rotation eucalyptus plantations in Portugal to complement an existing whole stand model, namely the GLOBULUS model presented in (Tomé et al. 2001). To achieve this objective, the following steps were needed:

1. Select a probability density function appropriate to model the diameter distribution of first-rotation eucalyptus plantations. Accordingly, at the start of the research and based on previous research (Furtado 2006), we tested the hypothesis that the empirical diameter distributions followed a Johnson's  $S_B$  distribution.
2. Test and implement an algorithm for the estimation of the parameters of the selected probability density function through the parameter recovery method. This method, not requiring the availability of a sample of diameters, is appropriate to estimate the probability density function parameters from stand variables. The algorithm proposed by Parresol (2003) was selected as a starting point for the parameter recovery.
3. Design a methodology to estimate the stand variables required for the parameter recovery method so that the diameter distribution can be simulated when only stand variables are available.
4. Evaluate the model developed.

## 2 Methods

### 2.1 Data

Permanent plots installed in 1970 and measured at 1-year interval were used to model diameter distributions of eucalyptus (*E. globulus* Labill.) plantations in Portugal. Therefore, the information concerning all trees (i.e. with no diameter threshold) includes successive measurements, usually annually, of the diameter at breast height ( $d$ ).

The data used to evaluate the model developed in this study were selected from the forest inventory of the Silvicaima Pulp and Paper Company. In this data set, trees with a diameter at breast height smaller than 5 cm were not measured, but they were counted.

Table 1 summarises data used to build and evaluate the model.

2.2 Testing the performance of Johnson’s  $S_B$  distribution

The analysis of the coefficients of skewness ( $\sqrt{\beta_1}$ ) and kurtosis ( $\beta_2$ ) allows the selection of the distribution followed by a certain population without the need to estimate any parameters. Figure 1 shows the location of the Weibull, normal, log-normal and various types of Johnson’s distributions in the plane  $(\beta_1, \beta_2)$ . Other studies have based the selection of a distribution on goodness of fit tests (Cao 2004; Zhang et al. 2003). The selection of the probability density function by this procedure may be dependent on the methodology used to estimate the parameters, not allowing the selection of the best density function which depends only on the data.

To first identify the distribution that better reproduced the set of observed frequencies, the estimates of

the coefficients of skewness ( $\sqrt{\beta_1}$ ) and kurtosis ( $\beta_2$ ) in each plot were first analysed. Below are the estimators used for skewness and kurtosis, respectively:

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad \text{and} \quad b_2 = \frac{m_4}{m_2^2}$$

with

$$m_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^j}{n - 1} \quad \text{and} \quad m_j = \frac{\sum_{i=1}^n (x_i - \bar{x})^j}{n}, \quad j = 3, 4.$$

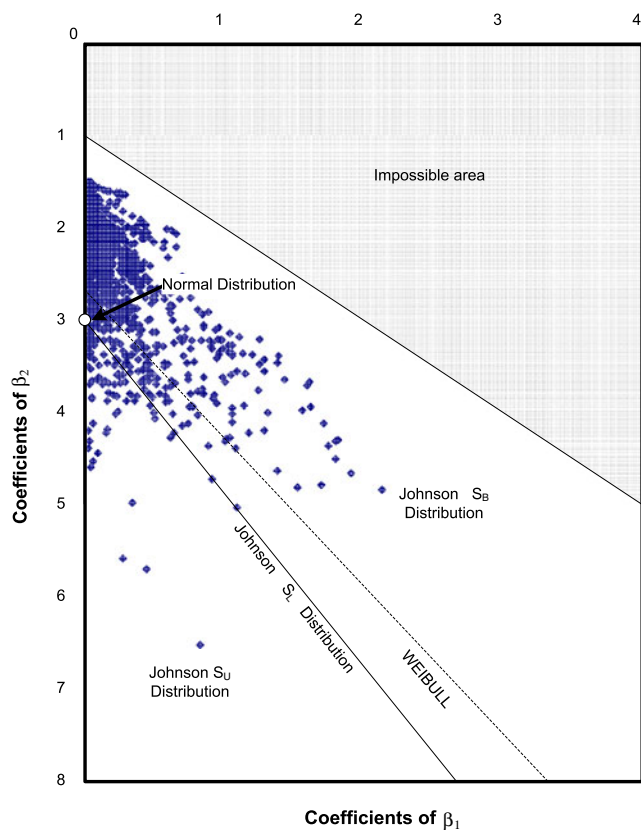
The choice of the Johnson’s  $S_B$  distribution as the null hypothesis for modelling the diameter distributions of eucalyptus was based on its flexibility to model distributions with different shapes. It has a broader range over the  $(\beta_1, \beta_2)$  space than other distributions and includes most of the alternative probability density functions (Johnson 1949; Johnson and Kotz 1970).

Since Hafley and Schreuder (1977) introduced the four-parameter Johnson’s  $S_B$  distribution into the forest literature, this probability density function has been

**Table 1** The number of measurements used by site index (base age 10), number of trees per hectare (at the time of planting) and age (years)

Site index class (m)	No of trees./ha classes	Age class (years)				Total
		≤4	4–8	8–12	≥12	
≤12	500–1,000	2				2
	1,000–1,500		1	3	11	15
	1,500–2,000					
	2,000–2,500			2	6	8
	2,500–3,000					
12–16	500–1,000	1	1 (2)	3 (14)	4 (5)	9 (21)
	1,000–1,500	6	34	43 (4)	24 (3)	107 (7)
	1,500–2,000	1	3	7 (2)	6	17 (2)
	2,000–2,500		3	3 (1)		6 (1)
	2,500–3,000					
16–20	500–1,000	5	4 (3)	12 (24)	5 (36)	26 (63)
	1,000–1,500	49	89 (9)	61 (15)	16 (29)	215 (53)
	1,500–2,000	4	9 (8)	6 (4)		19 (12)
	2,000–2,500	5	7 (5)	7 (1)	5	24 (6)
	2,500–3,000		3 (1)	11	7	21 (1)
20–24	500–1,000	11	19 (2)	15 (10)	5 (44)	50 (56)
	1,000–1,500	41	59 (12)	65 (20)	61 (23)	226 (55)
	1,500–2,000	18	23 (13)	14 (8)	17	72 (21)
	2,000–2,500	11	8 (5)	3	6	28 (5)
	2,500–3,000	6	6 (1)	5	31	48 (1)
24–28	500–1,000	5	5	5 (2)	3 (1)	18 (3)
	1,000–1,500	15	13 (1)	21 (8)	22 (1)	71 (10)
	1,500–2,000	5	6 (4)	7 (3)		18 (7)
	2,000–2,500		4	10		14
	2,500–3,000			1	10	11
>28	500–1,000					
	1,000–1,500	1	3		7	11
	1,500–2,000			3	1	4
	2,000–2,500					
	2,500–3,000					
						1,040 (324)

The number of measurements used for validation are in round brackets



**Fig. 1** Region in the plane of skewness ( $\beta_1$ ) and kurtosis ( $\beta_2$ ) coefficients that corresponds to the normal, Weibull and different types of Johnson’s distributions (Hahn and Shapiro 1967)

widely used in forest diameter (and height) distribution modelling by several authors (Hafley and Buford 1985; Fonseca 2004; Fonseca et al. 2009; Kamziah et al. 1999; Li et al. 2002; Parresol 2003; Rennolls and Wang 2005; Zhang et al. 2003; Zhou and McTague 1996). Furtado (2006) compared several probability density functions to model eucalyptus diameter distribution, including the Weibull, log-normal, gama, beta and Johnson’s system and concluded that the Johnson’s  $S_B$  distribution fits well to the distribution of diameters in eucalyptus populations in Portugal. To confirm the performance of the Johnson’s  $S_B$  distribution in modelling the diameter distributions of eucalyptus plantations in Portugal, the  $b_1$  and  $b_2$  estimates were computed for each measurement on the fitting data set and plotted over the plane ( $\beta_1, \beta_2$ ) to check if the pairs ( $b_1, b_2$ ) occurred mainly in the area that corresponds to this distribution.

### 2.3 The Johnson system of probability density functions

The Johnson system corresponds to the distribution of a random variable ( $X$ ) in which a particular trans-

formation is applied to obtain a normal distribution. This system is composed of three kinds of distributions (i.e. the Johnson’s  $S_L, S_B$  and  $S_U$ ) that depend on the transformation applied to the random variable (Johnson 1949).

When the transformation  $Z = \gamma + \delta g(X)$  is applied to the random variable ( $X$ ), an infinite system of distribution functions (or random variables) is defined according to the transformation  $g(X)$ , which is necessary to obtain a transformation with standard normal distribution.

Johnson (1949) introduced four parameters, namely  $\gamma, \delta, \epsilon$  and  $\lambda$ , with  $\gamma, \epsilon \in \mathbb{R}, \lambda \in \mathbb{R}^+, \delta \in \mathbb{R}/\{0\}$  and expressed the generic transformation defined above in the following way:

$$Z = \gamma + \delta g\left(\frac{X - \epsilon}{\lambda}\right), \tag{2.1}$$

where  $\gamma$  and  $\delta$  are shape parameters and  $\epsilon$  and  $\lambda$  are location and scale parameters, respectively. Although parameters  $\gamma$  and  $\delta$  affect both the skewness and the kurtosis of the distribution, parameter  $\gamma$  is particularly associated with asymmetry, and an increase in parameter  $\delta$  corresponds to an increase in the kurtosis (Johnson 1949).

To generate distributions with limited domain, the following transformation was chosen:

$$g(Y) = \ln\left(\frac{Y}{1 - Y}\right). \tag{2.2}$$

In terms of  $Y = \frac{X - \epsilon}{\lambda}$ , this results in:

$$Z = \gamma + \delta \ln\left(\frac{X - \epsilon}{\epsilon + \lambda - X}\right), \quad \epsilon < X < \epsilon + \lambda, \\ -\infty < \gamma < \infty, \delta > 0, -\infty < \epsilon < \infty, \lambda > 0 \tag{2.3}$$

or

$$Z = \gamma + \delta \ln\left(\frac{Y}{1 - Y}\right), \quad 0 < y < 1, \\ -\infty < \gamma < \infty, \delta > 0, -\infty < \epsilon < \infty, \lambda > 0. \tag{2.4}$$

The system of random variables generated by Eqs. 2.3 and 2.4 is called the Johnson’s  $S_B$  system of distributions.

### 2.4 Estimating the parameters of the Johnson’s $S_B$ distribution from stand variables

The parameters of the Johnson’s  $S_B$  distribution were recovered using a methodology similar to the one proposed by Parresol (2003).

If Eq. 2.4 is expressed in terms of  $Y$ , the following expression is obtained:

$$Y = \left[ 1 + \exp\left(-\frac{Z - \gamma}{\delta}\right) \right]^{-1}. \tag{2.5}$$

When  $Z$  assumes the null value, the median of the variable  $Y$  is obtained:

$$y_{1/2} = (1 + e^{\gamma/\delta})^{-1}. \tag{2.6}$$

Note that the medians of  $Y$  and  $X$  are related because  $y_{1/2} = \frac{x_{1/2} - \epsilon}{\lambda}$ .

Note also that Eq. 2.6 enables the estimation of the shape parameter,  $\gamma$ , according to the median value of the diameter distribution, provided that the shape parameter  $\delta$  is known:

$$\gamma = \delta \ln\left(\frac{1}{y_{1/2}} - 1\right) = \delta \ln\left(\frac{\lambda}{x_{1/2} - \epsilon} - 1\right). \tag{2.7}$$

However, another equation is needed to estimate the shape parameter  $\delta$ . In the growth of trees, competition between trees affects growth in tree diameter. Thus, a variable of great importance in the elaboration of stand models using diameter distribution simulation is the stand basal area ( $G$ ). This variable is related to the second moment not centred ( $E(X^2)$ ) on the distribution:

$$G = \frac{1}{10,000} \sum_{i=1}^N \frac{\pi}{4} d_i^2 = c N E(X^2) \quad (\text{m}^2 \text{ ha}^{-1}), \tag{2.8}$$

where  $N$  is number of trees per hectare and  $d_i$  is diameter at breast height (centimetres) of tree  $i$  and  $c = \frac{\pi}{40,000}$  is a conversion constant.

As

$$E(X^2) = E(\epsilon + \lambda Y)^2 = \epsilon^2 + 2\epsilon\lambda E(Y) + \lambda^2 E(Y^2), \tag{2.9}$$

then

$$G = c N (\epsilon^2 + 2\epsilon\lambda E(Y) + \lambda^2 E(Y^2)). \tag{2.10}$$

The non-centered moments of order  $r$  ( $E(Y^r)$ ) may be determined through the moment-generating function  $\varphi$  of the variable  $Y$ :

$$\varphi_Y(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{t}{1 + e^{-\frac{z-\gamma}{\delta}}}\right) e^{-z^2/2} dz$$

which shows the following relationship:

$$E(Y^r) = \varphi_Y^{(r)}(t) |_{t=0} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(1 + e^{-\frac{z-\gamma}{\delta}}\right)^{-r} e^{-z^2/2} dz.$$

By assuming some reasonable values for  $\epsilon$  and  $\lambda$ , the solution of the system comprised of Eqs. 2.7 and 2.10 and based on known values of  $y_{1/2}$ ,  $G$  and  $N$  allows for the estimation of parameters  $\gamma$  and  $\delta$ . The solution requires the use of iterative methods for numerical integration, as the calculation of the moments of the distribution does not yield an analytical solution.

This algorithm was implemented using Visual Basic under the SIMFLOR interface that can be easily added to existing growth and yield models (Marto et al. 2009).

As in any iterative process, it is necessary to assign initial values to the parameters. Parresol (2003) suggests using the initial value for  $\delta$  to estimate an initial value for  $\gamma$  according to Eq. 2.7. Parameter  $\delta$  was initialised with the estimate obtained for the kurtosis because an increase in  $\delta$  corresponds to an increase in kurtosis (Johnson 1949). Parameter  $\epsilon$  was set equal to the minimum value of the observed diameter, and  $\lambda$  was initialised using the difference between the maximum and minimum value of the observed diameters. The values for  $G$  and  $N$  were obtained from measurements in the study plots.

To complement the methodology used, based on the analysis of coefficients  $(\sqrt{\beta_1}, \beta_2)$  for a first identification of the distribution to be used, the goodness-of-fit Kolmogorov–Smirnov test was also used in order to test the hypothesis that the Johnson’s  $S_B$  distribution fitted to the diameter distributions on individual plots (Law and Kelton 1982; Massey 1951; Reynolds et al. 1988). We used the modified Kolmogorov–Smirnov test because the parameters were unknown and estimated from the data (Lilliefors 1967). The chi-squared test was not used because of dependency resulting from the grouping of data in classes.

## 2.5 Modelling the diameter distributions of eucalyptus

### 2.5.1 The estimation of the stand variables required to fit the distribution

Modelling of the diameter distributions of eucalyptus was based on the development of a system of equations that complement the output of an existing whole stand growth model (GLOBULUS, Tomé et al. 2001) by estimating the stand variables required to fit the distribution, including the average ( $\bar{d}$ ), median ( $d_{1/2}$ ), minimum ( $d_{\min}$ ) and maximum diameter ( $d_{\max}$ ) of the diameter distribution. Other stand variables used in the system of equations were taken from the output of the GLOBULUS model (Tomé et al. 2001), including the site index ( $S$ ), dominant height ( $h_{\text{dom}}$ ), stand density ( $N$ ) and basal area ( $G$ ).

The development of the models to estimate  $\bar{d}$ ,  $d_{1/2}$ ,  $d_{\min}$  and  $d_{\max}$  was based on the presumed relationships between stand variables (e.g.  $N$ ,  $S$ ,  $G$ ,  $d_g$ ,  $h_{\text{dom}}$ ,  $t$ ,  $\bar{d}$ ,  $d_{1/2}$ ,  $d_{\min}$ ). In fact, some restrictions were imposed on the construction of the model to obtain estimates that observed well-established biological processes. Thus, the minimum diameter ( $d_{\min}$ ) should be lower than the median value ( $d_{1/2}$ ) and the average diameter ( $\bar{d}$ ). Additionally, average diameter ( $\bar{d}$ ) should be lower than the quadratic mean diameter ( $d_g$ ) such that

$$d_{\min} < \min(\bar{d}, d_{1/2})$$

and

$$\bar{d} < d_g.$$

The quadratic mean diameter in centimetres was obtained from the relationship

$$d_g = 100 \sqrt{\frac{4 G}{\pi N}}.$$

In each of the four models developed, the independent variables that explained the greatest proportion of the variability of the dependent variable were selected. For the mean diameter, the model was formulated such that

$$\bar{d} = d_g - (\text{proportion of } d_g).$$

In this way, the properties  $\bar{d} < d_g$  and  $\bar{d} > 0$  were guaranteed. The proportion of the quadratic mean diameter (proportion of  $d_g$ ) was modelled using the function  $1/(1 + \exp(f))$ , where  $f$  is a function of stand variables:

$$\bar{d} = d_g - \frac{1}{1 + \exp(f)} d_g.$$

The median diameter ( $d_{1/2}$ ) was modelled as a multiple allometric relationship between the stand characteristics. The minimum diameter ( $d_{\min}$ ) was formulated in a way similar to the mean diameter while ensuring that  $d_{\min} < \min(\bar{d}, d_{1/2})$  and  $d_{\min} > 0$ :

$$d_{\min} = \min(\bar{d}, d_{1/2}) - \frac{1}{1 + \exp(g)} \min(\bar{d}, d_{1/2}),$$

where  $g$  is a function of stand variables.

Finally, ( $d_{\max}$ ) was simulated through a multiple allometric relationship. One of the variables used in this simulation was the variance of the diameters (vard), which was calculated from the relationship  $V(X) = E(X^2) - E^2(X)$  so that:

$$\text{vard} = \frac{4}{\pi} G N - (\bar{d})^2.$$

As a starting point, the models were individually fitted to select the variables to be included as regressors in each of the models. The homoscedasticity was analysed using the plot of studentised residuals versus predicted values, while the normality of model errors was inspected using a normal probability plot. A weight regression and/or the Huber function for robust regression was used when heteroscedasticity or non-normality of the errors was present, respectively (Myers 1990).

The selection of the regressors was based on their predictive abilities evaluated using the PRESS residuals (Myers 1990). Bias was evaluated using the average press residual, and precision was estimated by the absolute value of the residuals as well as by the 5th and 95th percentiles. Modelling efficiency (or the percentage of variance explained by the model) was also computed with the PRESS residuals.

With the exception of  $G$ ,  $N$  and  $h_{\text{dom}}$ , which were taken from an existing model (GLOBULUS, Tomé et al. 2001), the equations selected for each of the variables in the diameter distribution model were simultaneously fit to minimise the determinant of the covariance matrix of errors. A simultaneous fit was needed because in this set of functions, the independent variables appear in more than one function, and dependent variables of some functions appear as independent variables in other functions. This adjustment was made using the three-stage method of least squares (three-stage least squares) (Gallant 1987, pp 427–443). The PROC MODEL-Statement from (SAS 2005) was used. During the simultaneous fitting, each of the component equations was weighted as in the individual fitting, and the weights calculated with the Huber function were included (Myers 1990).

**Table 2** Evaluation of the fit and predictive ability of the models

Model	Average $r$	Average $ r $	Percentile 5%	Percentile 95%	Modelling efficiency
Average ( $\bar{d}$ )	-0.03	0.26	-0.66	0.42	0.9912
Median ( $d_{1/2}$ )	-0.01	0.39	-0.89	0.63	0.9811
Minimum diameter ( $d_{\min}$ )	-0.11	1.22	-2.30	3.22	0.4421
Maximum diameter ( $d_{\max}$ )	0.18	1.39	-2.65	3.82	0.9412

$r$  PRESS residual

**Table 3** System of equations used to estimate the stand attributes that were, in turn, used to recover the Johnson’s  $S_B$  distribution

Model	Analytical expression
Number of trees per ha ( $N$ , ha <sup>-1</sup> ) (Tomé et al. 2001)	$N_i = N_{i-1}e^{-a(t_i-t_{i-1})}$ $a = f$ (site index, density, climatic region)
Dominant height ( $h_{dom}$ , m) (Tomé et al. 2001)	$h_{domi} = A_h \left( \frac{h_{domj}}{A_h} \right) \left( \frac{t_j}{t_i} \right)^n$ $n = f$ (rotation, climatic region)
Basal area ( $G$ , m <sup>2</sup> ha <sup>-1</sup> ) (Tomé et al. 2001)	$G_i = A_g \left( \frac{G_{i-1}}{A_g} \right) \left( \frac{t_{i-1}^{ng1}}{t_i^{ng2}} \right)$ $n_{1g} e n_{2g} = f$ (rotation, climatic region, $N_i$ or $N_{i-1}$ respectively)
Average ( $\bar{d}$ , cm)	$\bar{d} = d_g - \frac{1}{1 + \exp(-0.2508 h_{dom} + 0.0829 S + 0.000746 N + 0.4058 d_g)} d_g$
Median ( $d_{1/2}$ , cm)	$d_{1/2} = 1.2016 d_g^{1.1071} h_{dom}^{-0.0931} N^{-0.0293}$
Minimum diameter ( $d_{min}$ , cm)	$d_{min} = \min(\bar{d}, d_{1/2}) - \frac{1}{1 + \exp(-0.00081 N - 0.0367 h_{dom} 0.0407 \bar{d})} \min(\bar{d}, d_{1/2})$
Maximum diameter ( $d_{max}$ , cm)	$d_{max} = 2.3483 \text{ vard}^{0.1342} d_g^{0.4806} h_{dom}^{0.2008}$

$a, n, ng1, ng2, A_h$  and  $A_g$  are parameters dependent on the climatic region (see Tomé et al. 2001)  
 $N_i$  is  $N$  at time  $t_i$ ,  $h_{domi}$  is  $h_{dom}$  at time  $t_i$ ,  $d_g$  is quadratic mean diameter,  $G_i$  is  $G$  at time  $t_i$

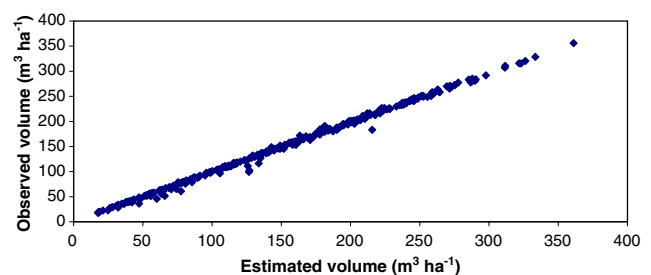
2.5.2 Validation of the recovered model

The system of equations developed above was used to predict the average, median, minimum and maximum diameters for each measurement in each of the plots available for validation. The respective diameter distribution was then recovered using the procedure described above. Based on the study of the capacity of the Johnson’s  $S_B$  to simulate the distribution of eucalyptus diameters, the values for the kurtosis were approximately equal to 2. Thus, an initial value of 2 was always used. An important question in validating a model is not whether the model is correct but rather how close predictions from the model are to what is observed in empirical forest stands. Goodness-of-fit tests were not used to validate the model because these tests are not designed to address issues regarding the error associated with estimates and predictions from this type of model.

Our main objective after modelling the diameter distributions of a stand was to evaluate the size of the error that may result for the difference between predicted total volume and observed total volume. To achieve this objective, we used an error index as an overall measure of model performance by putting all diameters into one class (Reynolds et al. 1988). The statistical methods used to measure model performance included the  $t$  test and the linear regression between

observed and estimated volume. In summary, two types of tests were conducted using the data available for validation:

1. A paired sample  $t$  test was used to assess if the volume estimated for each plot in each age using the simulated diameter distribution was significantly different from the corresponding observed volume (Reynolds et al. 1988).
2. Observed volumes ( $V$ ) were also plotted against the corresponding estimated values ( $\hat{V}$ ), and a simple linear regression model  $V = a + b \hat{V} + \epsilon$  was fitted to the data, where  $a + b \hat{V}$  is the deterministic component of the model and  $\epsilon$  is the associated stochastic error.



**Fig. 2** The volume observed in each plot at each age available in the validation data set versus the volume estimated with the simulated diameter distribution



### 3 Results

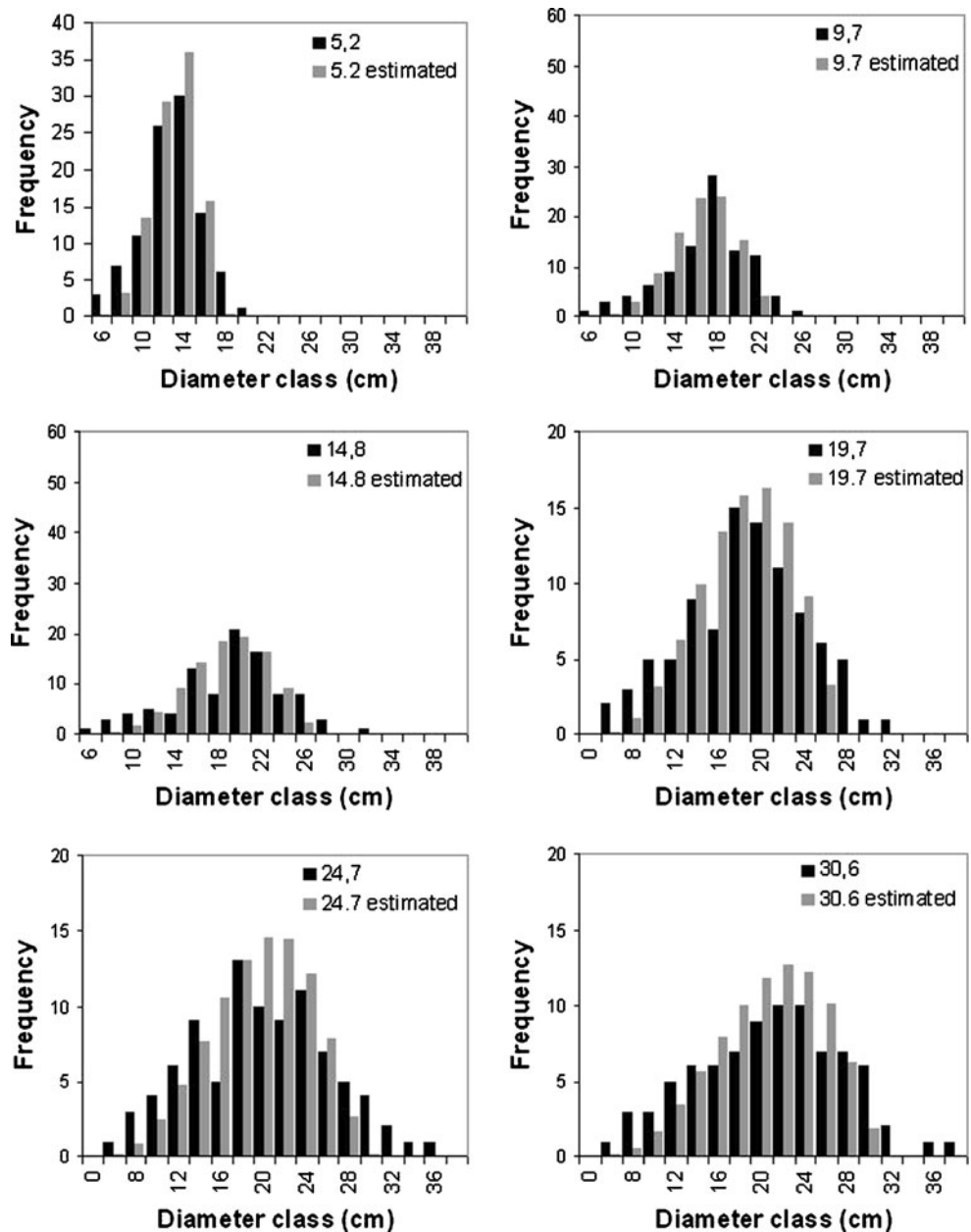
#### 3.1 The fit of the Johnson's $S_B$ distribution to observed diameter distributions

The ranges for the coefficients  $(\sqrt{b_1}, b_2)$  estimated with the data set described in Table 1 were  $-1.3977 \leq \sqrt{b_1} \leq 1.0805$  and  $1.8112 \leq b_2 \leq 6.8685$ , which indicates the existence of a huge variety of empirical diameter distributions for eucalyptus plantations. This supports the choice of a very flexible distribution. Figure 1 illustrates the location of the pairs  $(b_1, b_2)$ ; it can be seen that most of the values observed for

the plots under study are included in the Johnson's  $S_B$  region. This supports the hypothesis presented in the introduction.

It was also verified that in the great majority of the plots, the coefficients of skewness assumed negative values. In the growth of trees, competition between trees affects growth in tree diameter; this fact explains the negative values for the coefficients of skewness. In other words, the trees that had a small initial growth in diameter ( $d$ ) continued to have low growth rates, and the differences between small and large trees tended to increase (Perry 1985; Soares and Tomé 1996). The modified Kolmogorov–Smirnov test showed that the

**Fig. 3** Comparison of real and estimated diameter distribution from a plot in first rotation at ages (years) 5.2, 9.7, 14.8, 19.7, 24.7 and 30.6 (dark real values)



Johnson's  $S_B$  distribution did not significantly differ from the empirical distribution in 106 out of 111 studied stands in which each had several measurements for tree stands aged 5 to 32 years.

### 3.2 Models needed for parameter recovery

#### 3.2.1 The prediction of stand variables

Table 2 summarises the statistics used to characterise the predictive capacity of individual models. Table 3 shows the final models in which parameters were simultaneously fitted to estimate the stand attributes such that stand attributes were used to recover the diameter distribution. With the exception of  $d_{\min}$ , all models had a very high modelling efficiency and a well predictive ability with bias less than 2 mm and precision between 2 and 15 mm. The relatively low modelling efficiency of  $d_{\min}$  was mainly due to the small total variability of this variable, which therefore does not indicate low predictive power. The final system (see Table 3) includes equations that are all consistent with well-established biological principles.

#### 3.2.2 Johnson $S_B$ parameter estimation and validation

The models shown in Table 3 were used to recover the Johnson's  $S_B$  parameters. Parameter  $\delta$  was always initialised with a value equal to 2. Figure 2 shows that the volume observed in each plot for each age is very close to the volume estimated using the diameter distribution. The paired sample  $t$  test showed that there were no significant differences between observed and estimated volumes at the 5% significance level. Based on the results from the adjusted regression, the coefficient  $b$  did not differ from 1 at the 5% significance level, thus indicating a well agreement between the observed and predicted volumes.

As an example, Fig. 3 shows the evolution of the observed and simulated diameter distribution from 5.2 to 30.6 years of age in one of the permanent plots from the fitting data set when the initialisation was made with the measurement made at 5.2 years of age. As can be seen, the agreement is very good, even for ages far away from the initial one. The results in other long-term series plots were similar.

## 4 Discussion

The objective of the research described here was to model diameter distributions of *E. globulus* plantations

in Portugal and to develop a methodology to model the diameter distributions from stand characteristics that can be obtained from an existing growth model and/or from other prediction functions. As with several other species (Hafley and Buford 1985; Fonseca 2004; Kamziah et al. 1999; Li et al. 2002; Parresol 2003; Zhang et al. 2003; Zhou and McTague 1996), the Johnson's  $S_B$  probability density function, which is very flexible in terms of the forms that it can take, performed very well.

A preliminary study on the distribution of diameters in eucalyptus populations in Portugal concluded that in general, the Johnson's  $S_B$  distribution adequately approximates stands in which the site index is above a given level of productivity, namely 12 m at age 10, which characterises the large majority of industrial plantations in the country (Furtado 2006). These previous findings are strengthened by the present research, which has drawn on a very large data set covering stands with different characteristics (see Table 1).

As can be seen in Fig. 1, the distribution of the pairs  $(b_1, b_2)$  covers a large area in the plane of  $(\beta_1, \beta_2)$ , which underscores the need for a flexible probability density function. Figure 1 also shows that alternative functions used by other authors, such as the Weibull functions (Páscoa 1987; Palahi et al. 2007), would not have been able to model such large variation. Importantly, the present research suggests that a study of the distribution of the pairs  $(b_1, b_2)$  in the plane  $(\beta_1, \beta_2)$  must be undertaken before selecting candidate probability distributions functions for modelling the diameter distributions of forest populations.

Palahi et al. (2007) compared the performance of different probability functions as diameter distributions of forest stands in Catalonia, and they found that the truncated Weibull function appeared to be most accurate and consistent for characterising the distribution of stand basal areas, performing even better than the Johnson's  $S_B$  distribution. According to Fig. 1, this result would not be expected. Although the Weibull distribution is easy to adjust, it is less flexible than the Johnson's  $S_B$ . In fact, it is represented in the  $(\beta_1, \beta_2)$  plane by a line corresponding to a sub-space of the area covered by the Johnson's  $S_B$  distribution (see Fig. 1). Accordingly, the Johnson's  $S_B$  would be expected to perform better. However, the methodology utilised by these authors had some limitations because the sampling method was not specifically designed to develop models and estimate diameter distributions. Small trees (<7.5 cm) were not measured, and in addition, most of the plots had too few trees to characterise the diameter distribution of the whole stand. These limitations could account for this unexpected result. In the present

work, after based on the range of values in  $(\sqrt{b_1}, b_2)$  and the Kolmogorov–Smirnov modified test, we determined that the Johnson's  $S_B$  distribution was the most appropriate probability density function to model the distribution of tree diameters in eucalyptus plantations, and the results of the present research support this decision.

Once the Johnson's  $S_B$  distribution was selected, an algorithm was implemented to recover its parameters from the predicted values of stand variables using a Visual Basic program (Marto et al. 2009). The method closely followed the algorithm proposed by Parresol (2003). To implement the proposed algorithm, the following stand variables must be predicted, namely basal area, dominant height, number of trees per hectare, average diameter, median diameter, minimum diameter and maximum diameter. The first three variables could be obtained from an existing whole stand model, that is, the GLOBULUS model (Tomé et al. 2001); the last four variables can be predicted using the system of equations developed in this work (Table 3). The equations were designed by taking into account the relationships between stand variables to obtain estimates that were in line with well-established biological principles. Parresol (2003) also imposed restrictions on stand equations to ensure that the parameters made biological sense. In fact, when modelling  $\bar{d}$ , this author used the logarithm of  $d_g - \bar{d}$  as the dependent variable to account for the fact that the average diameter should be lower than the quadratic mean diameter. The approach used in the present work is preferable because the average diameter is directly used as the dependent variable without the need for any transformation, which therefore leads to better predictions. Additionally, the average diameter was lower than the quadratic mean diameter.

The main advantage of using a parameter recovery model is that with the stand variables in parameter recovery, compatibility between the characteristics of the observed population and those obtained through the theoretical distribution is assured. Fonseca et al. (2009) uses the parameter recovery method to recover all four parameters of Johnson's  $S_B$  distribution; however, the method developed by these authors requires stand variables for future applications that are not easy to model, and the achieved improvement is not substantial. Rennolls and Wang (2005) presented a new parameterisation of the Johnson's  $S_B$  distribution using maximum likelihood estimation. However, it requires data on diameters to fit the distribution and, consequently, cannot be used in growth and yield models, which is the objective of this research.

## 5 Conclusions

In conclusion, this study aimed at developing a model to simulate diameter distributions for eucalyptus plantations in Portugal revealed the following:

- The Johnson's  $S_B$  distribution is appropriate to simulate diameter distributions of *Eucalyptus* plantations in Portugal.
- The parameter recovery method, assuring compatibility between the stand characteristics stand density and basal area and the same stand characteristics estimated through the Johnson's  $S_B$  distribution, originates volume estimations close to the volumes estimated from the observed individual trees (the best possible estimate).
- The system of equations provided in this research includes new regression models for average, median, minimum and maximum diameter distribution that were built in a way to obtain estimates that observed logical relationships according to biological principles.
- The system of equations provided in this paper are an important added value to the GLOBULUS model and allow the model to yield better information for the computation of wood assortments and for planning harvesting operations.

In this paper, methods were presented to select a probability density function, and a model was developed to simulate the diameter distribution when only stand variables are available. These methods are general and can be used for other species and regions, provided that the system of equations presented in Table 3 is refit with appropriate data.

**Acknowledgements** This work was supported by the FP6 EFORWOOD IP project (contract 518128) and the FCT project CarbWoodCork (POCI/AGR/57279/2004). The authors gratefully acknowledge Bernard Parresol (USDA Forest Service Southern Research Station) for facilitating the use of a SAS program for the Johnson's  $S_B$  distribution parameter recovery that was used in an initial stage of this research. The authors also thank the pulp companies Celbi and Silvicaima for providing a large portion of the data used in this study.

## References

- Aranda UD (2004) Modelo dinámico de crecimiento para masas de *Pinus sylvestris* L. procedentes de repoblación en

- Galicia. Tesis Doctoral, Universidade de Santiago de Compostela -Escuela Politécnica Superior-Depart. de Ingeniería Agroflorestal, 310 p
- Bailey RL, Dell TR (1973) Quantifying diameter distributions with the Weibull function. *For Sci* 19(2):97–104
- Cao QV (2004) Predicting parameters of a Weibull function for modeling diameter distribution. *For Sci* 50(5):682–685
- Fonseca TF (2004) Modelação do crescimento, mortalidade e distribuição diamétrica, do pinhal bravo no vale do Tâmega. Tese de Doutoramento, Universidade de Trás-os-Montes e Alto Douro. 247 p.
- Fonseca TF, Marques CP, Parresol BR (2009) Describing maritime pine diameter distributions with Johnson's  $S_B$  distribution using a new all-parameter recovery approach. *For Sci* 55(4):367–373
- Furtado AX (2006) Modelação da estrutura dinâmica de povoamentos de Eucalyptus globulus em primeira rotação. Tese de Doutoramento, Faculdade de Ciências e tecnologia da Universidade Nova de Lisboa, 192 p
- Gallant A (1987) Nonlinear statistical models. Wiley, New York, 624 p
- Hafley WL, Buford MA (1985) A bivariate model for growth and yield prediction. *For Sci* 31:237–247
- Hafley WL, Schreuder HT (1977) Statistical distributions for fitting diameter and height data in even-ages stands. *Can J For Res* 7:481–487
- Hahn GJ, Shapiro SS (1967) Statistics for engineers. Wiley, New York, p 200
- Hyink DM, Moser Jr JW (1983) A generalised framework for projecting forest yield and stand structure using diameter distributions. *For Sci* 29(1):85–95
- Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36:147–176
- Johnson N, Kotz S (1970) Continuous univariate distribution, vol 1. Wiley, New York, 761 p
- Kamziah AK, Ahmad MI, Lapongan J (1999) Nonlinear regression approach to estimating Johnson SB parameters for diameter data. *Can J For Res* 29:310–314
- Kiviste A, Nilson A, Hordo M, Merenakk M (2003) Diameter distribution models and height-diameter equations for Estonian forest. In: Amaro A, Reed D, Soares P (eds) Modelling forest systems. CABI, Wallingford, pp 169–179
- Law AM, Kelton WD (1982) Simulation modelling and analysis. McGraw-Hill, New York, 400 p
- Li F, Zhang L, Davis CJ (2002) Modelling the joint distribution of tree diameters and heights by bivariate generalised beta distribution. *For Sci* 48(1):47–58
- Lilliefors HW (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Am Stat Assoc J* 62:399–402
- Maltamo M, Puumalainen J, Paivinen R (1995) Comparison of Beta and Weibull functions for modelling basal areas diameter distributions in stands of *Pinus sylvestris* and *Picea abies*. *Scand J For Res* 10:284–295
- Marto M, Palma J, Mateus A, Tomé M (2009) Computer program for estimation of Johnson's  $S_B$  parameters using a parameter recovery method. *Publicações Científicas Forchange PC-X/2009*. Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa
- Massey FJ (1951) The Kolmogorov–Smirnov test for goodness of fit. *Am Stat Assoc J* 46:68–78
- Ek AR, Monserud RA (1979) Performance and comparisons of stand growth models based on individual tree and diameter class growth. *Can J For Res* 9:231–244
- Myers RH (1990) Classical and modern regression with applications, 2nd edn. Duxbury, Belmont, 488 p
- Palahí M, Pukkala T, Blasco E, Trasobares A (2007) A comparison of beta, Johnsons SB, Weibull and truncated Weibull functions for modelling the diameter distribution of forest stands in Catalonia (north-east of Spain). *Eur J For Res* 126(4):563–571
- Parresol B (2003) Recovering parameters of johnson's  $S_B$  distribution. Res. pap. SRS-31. USDA For. Ser., Southern Research Station, Asheville, p 9
- Páscoa F (1987) Estrutura, Crescimento e Produção em Povoamentos de Pinheiro Bravo. Um Modelo de Simulação. Tese de Doutoramento, Universidade Técnica de Lisboa—Instituto Superior de Agronomia
- Perry DA (1985) The competition process in forest stands. In: Cannel MGR, Jackson JE (eds) Attributes of trees as crop plants. Institute of Terrestrial Ecology, Abbots Ripton, pp 481–506
- Rennolls K, Wang M (2005) A new parameterisation of Johnson's  $S_B$  distribution with application to fitting forest tree diameter data. *Can J For Res* 35:575–579
- Reynolds MR, Burk TE, Huang W (1988) Goodness-of-fit tests and model selection procedures for diameter distributions models *For Sci* 34:377–399
- SAS Institute (2005) The SAS System for windows, 9.1 edn. SAS Institute, Cary
- Soares P, Tomé M (1996) Changes in eucalypt plantations structure, variability and relative growth pattern under different intraspecific competition gradients. In: Skovsgaard JP, Johannsen VK (eds) Modelling regeneration success and early growth of forest stands. Proceedings from the IUFRO conference, Copenhagen. Danish Forest and Landscape Research Institute, Hørsholm, pp 270–284
- Tomé M, Ribeiro F, Soares P (2001) O modelo GLOBULUS 2.1-Relatórios técnico científicos do GIMREF no 1/2001. Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa
- Tomé M, Barreiro S, Cortiçada A, Meyer A, Ramos T, Malico P (2007) Inventário florestal 2005-2006. Áreas, volumes e biomassas dos povoamentos florestais. Resultados Nacionais e por Nut's II e III. Publicações GIMREF RT 5/2007. Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa
- Zhang L, Packard KC, Liu C (2003) A comparison of estimation methods for fitting Weibull and Johnson's SB distribution to mixed spruce-fir stands in northeastern North America. *Can J For Res* 33(7):1340–1347
- Zhou B, McTague JP (1996) Comparison and evaluation of five methods of estimation of the Johnson system parameters. *Can J For Res* 26:928–935