Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
http://dx.doi.org/10.5424/fs/2014233-04939

Forest Systems 2014 23(3): 500-505
ISSN: 2171-5068
eISSN: 2171-9845

OPEN ACCESS

Short Communication. A comparison of estimation methods for fitting Weibull and Johnson's S_B functions to pedunculate oak (*Quercus robur*) and birch (*Betula pubescens*) stands in northwest Spain

Jose Javier Gorgoso-Varela[1]* and Alberto Rojo-Alboreca[2]

[1] Departamento de Biología de Organismos y Sistemas. Universidad de Oviedo. Escuela Politécnica de Mieres. C/ Gonzalo Gutiérrez Quirós, s/n. 33600 Mieres (Asturias), Spain. [2] Unidade de Xestión Forestal Sostible (UXFS). Departamento de Enxeñaría Agroforestal. Universidade de Santiago de Compostela. Escola Politécnica Superior. Campus Universitario, 27002 Lugo (Galicia), Spain

Abstract

Aim of study. In this study we compared the accuracy of the Weibull and the Johnson's S_B functions for describing diameter distributions in pedunculate oak (*Quercus robur* L.) and birch (*Betula pubescens* Ehrh.) stands.

Area of study. Galicia (Northwest Spain).

Material and Methods. A total of 172 diameter distributions in pedunculate oak and 202 in birch stands were finally evaluated. We compared the accuracy of three commonly used estimation methods of the Weibull and four estimation methods of the Johnson's S_B functions for describing these diameter distributions.

Main results. For *Quercus robur* L. stands, the most suitable methods were the Percentiles followed by Maximum Likelihood for the Weibull PDF and the method of Moments for the Johnson's S_B PDF. For *Betula pubescens* Ehrh. stands, the best fits obtained with the Percentiles and Maximum Likelihood methods were also superior to the method of Moments, whereas the Conditional Maximum Likelihood and method of Moments provided the best results for the Johnson's S_B PDF, depending on the statistic and the value of the location parameter considered.

Research highlights. Both distributions were suitable. The results were better for pedunculate oak than for birch stands.

Key words: Knoebel and Burkhart; location parameter; percentiles; maximum likelihood; moments, mode.

Introduction

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 purpose, they need tools to evaluate different management practices and their effects on stand structure. The development of growth models for the species (Rojo *et al.,* 2005; Diéguez-Aranda *et al.,* 2006; Gorgoso-Varela *et al.*, 2008) has enabled promotion of the productive and protective aspects of birch and pedunculate oak stands in northwest Spain. Diameter class models enable managers to predict stand growth and plan various uses, and they also provide data about stand structure. Such models are used to estimate stand variables and their structure with a probably density function (PDF) or a cumulative distribution function (CDF), either of which is fitted to diameter at breast height distributions or individual tree volume.

The main purpose of this study was to compare the accuracy of the Weibull and the Johnson's S_B PDFs, fitted with some of the most usual methods (3 methods for the Weibull PDF: Maximum Likelihood, Moments and Percentiles; and 4 methods for the Johnson's S_B PDF: Conditional Maximum Likelihood, Moments, Mode and Knoebel and Burkhart's method). The scope of the present paper is limited to fitting distributions to data comparing distributions with different numbers of parameters.

* Corresponding author: gorgoso@uniovi.es
Received: 03-09-13. Accepted: 07-07-14.

Table 1. Summary of main stand variables for the study stands

	Variable	Mean	Maximum	Minimum	Standard deviation
Quercus robur L. (N = 172)	d_g	21.9	40.0	8.6	6.1
	N	873.9	3,022.2	302.2	481.6
	H_0	17.0	25.6	7.2	3.3
	G	28.3	72.9	3.4	9.2
Betula pubescens Ehrh. (N = 202)	d_g	15.1	26.1	7.4	3.8
	N	1,690.9	6,000.0	350.0	1,077.6
	H_0	16.1	24.5	3.7	3.7
	G	26.7	76.4	3.3	11.1

d_g: Quadratic mean diameter (cm); N: Density (trees · ha^{-1}); H_0: Dominant height (m); G: Basal area ($m^2 \cdot ha^{-1}$).

Material and methods

Data set

In total, 172 diameter distributions in pedunculate oak and 202 in birch stands were finally evaluated in the present study. The size of the plots ranged from 200 m^2 to 1,000 m^2 in birch stands. In the pedunculate oak stands, plots ranged in size from 225 m^2 to 1,345 m^2 (depending on the stand density), to achieve a minimum of 30 trees per plot in both species.

Two perpendicular diameters at breast height were measured with calipers, to the nearest 0.1 cm, and the arithmetic average was calculated. The empirical data represent left-truncated distributions in many cases, as the smallest diameter measured in the field was 5 cm. In total, 16,210 diameter measurements in birch stands and 10,248 measurements in oak stands were available for analysis. Summary of main stand variables for the study stands are shown in Table 1.

Model fitting

The Weibull function

The three-parameter Weibull CDF is obtained by integrating the Weibull PDF, and has the following expression for a continuous random variable x:

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

where $F(x)$ is the cumulative relative frequency of trees with diameter equal to or smaller than x, a is the location parameter, b is the scale parameter and c is the shape parameter. Three methods of estimating the parameters of the Weibull distribution were compared: Percentiles (PW), Maximum Likelihood (ML) and Moments (MMW) used by Zhang *et al.* (2003).

Location parameter a of the function was considered in all fitting methods as d_{min} – c, with c = 5%, 10%, 15% and 20% of the minimum diameter observed in each distribution. Similar values proposed Zhang *et al.* (2003) if diameters of 10 cm are considered.

The Johnson's SB function

The model of the S_B PDF (Johnson, 1949) has the following expression for a continuous random variable x:

$$f(x) = \frac{\delta}{\sqrt{2\pi}} \cdot \frac{\lambda}{(\varepsilon + \lambda - x)(x - \varepsilon)} \cdot e^{-\frac{1}{2}\left[\gamma + \delta \cdot \ln\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right)\right]^2} \qquad [2]$$

where *f(x)* is the probability density associated with diameter x, $\varepsilon < x < \varepsilon + \lambda$, $-\infty < \varepsilon < +\infty$, $-\infty < \gamma < +\infty$, $\lambda > 0$, and $\delta > 0$.

The model is characterized by the location parameter ε, the scale parameter λ, and the shape parameters γ and δ (asymmetry and kurtosis parameters, respectively). Four methods of estimating the Johnson´s S_B parameters were compared: Conditional Maximum Likelihood (CML), Moments (MMJ), Mode (MJ) and Knoebel and Burkhart's (KB) method (Knoebel and Burkhart, 1991).

Location parameter ε of the function was considered in the CML, MMJ and MJ methods as d_{min} – c, with c = 5%, 10%, 15% and 20% of the minimum diameter observed in each distribution. In these fitting methods, the scale parameter λ of the function was considered as the maximum diameter observed in each distribution (d_{max}).

In the fits with the KB method, the location and scale parameters (ε and λ) were predetermined according to Knoebel and Burkhart (1991).

Model comparison

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

$$Bias = \frac{\sum_{i=1}^{N} Y_i - \hat{Y}_i}{N} \quad [3]$$

$$MAE = \frac{\sum_{i=1}^{N} \left| Y_i - \hat{Y}_i \right|}{N} \quad [4]$$

$$MSE = \frac{\sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2}{N} \quad [5]$$

where Y_i is the relative frequency of trees observed value in each diameter class, $\hat{Y}_i$ is the theoretical value predicted by the model, and N is the number of data points.

The Bias, MAE and MSE values were calculated for each fit in the mean relative frequency of trees for all combinations of diameter classes (1 cm) and plots. The Weibull PDF was used for reliable comparison of results instead the CDF.

The Kolmogorov-Smirnov (KS) statistic (D_n) for a given cumulative distribution function $F(x)$ was also used to evaluate and compare the results as Cao (2004): $D_n = \sup_x|F_n(x) - F_0(x)|$, where sup x is the supremum of the set of distances, where the cumulative observed frequency $F_n(x)$ is compared with the cumulative estimated frequency $F_0(x)$.

Results and discussion

The mean values of bias, mean absolute error (MAE), mean square error (MSE) in relative frequency of trees, and the mean value and the standard deviation of the Kolmogorov-Smirnov statistic (D_n) for the fits in *Quercus robur* stands (N = 172 plots) and the corresponding statistics for *Betula pubescens* stands (N = 202 plots) are shown in Table 2. The number of trees per ha observed and fitted by four methods for the Johnson's S_B PDF and three methods for the Weibull PDF, with c = 10% of the minimum observed diameter in four plots of *Quercus robur* and *Betula pubescens* stands, are shown in Fig. 1.

Both functions were suitable for fitting diameter distributions in pedunculate oak and birch stands in northwest Spain. The Johnson's S_B PDF provided the best results for the KS statistic (D_n) in both species except for MJ method, while the Weibull PDF generally provided the best fits, in terms of MAE and MSE. Bias may be less important in the comparison of results because errors with different signs can be compensated. The results were more accurate in pedunculate oak than in birch stands for all statistics compared.

For the fits of the Weibull PDF to *Quercus robur* L. stands, the most accurate results were obtained generally with the method of the Percentiles (PW), in terms of the MAE, MSE and D_n statistics and considering all the four parameters of location compared. However, the lowest value of the D_n statistic was obtained with the Maximum Likelihood (ML) approach, considering c = 5% of the minimum observed diameter. The most suitable value of c was 10% in terms of MSE in all three fitting methods. For the D_n statistic, the best results were obtained with c = 20% in PW, c = 5% in ML and c = 10% in MMW. Different values of the location parameter *a* of the Weibull PDF were computed in several studies (Río, 1999; Zhang *et al.,* 2003; Cao, 2004; Palahí *et al.,* 2007; Gorgoso *et al.,* 2012).

For the Johnson's S_B PDF, the lowest values of D_n were obtained with the Moments approach (MMJ), with c = 5% and c = 10%. The mode (MJ) method clearly provided the poorest results. Good fits with this method in same plots are shown in Fig. 1; however, in other cases the fits are clearly more biased than with the other methods (see the *B. pubescens* plot: 1BAR). Another problem with this method was in determining the mode value in same plots. Finally, the KB method was slightly inferior to MMJ and CML in terms of D_n, but values of MAE and MSE were similar to those obtained by the CML when c = 10%. Different values of the location parameter in Johnson's S_B function have been compared in several studies (Knoebel & Burkhart, 1991; Zhou & McTague, 1996; Zhang *et al.,* 2003; Scolforo *et al.,* 2003; Parresol, 2003; Palahí *et al.,* 2007; Fonseca *et al.,* 2009; Gorgoso *et al.,* 2012).

In the fits of the Weibull PDF to *Betula pubescens* Ehrh. stands, the most accurate results were obtained

Table 2. Mean values of bias, mean absolute error (MAE), mean square error (MSE) in relative frequencies of number of trees and mean value and standard deviation of the Kolmogorov-Smirnov (KS) statistic (D_n) for the fits with the Weibull and the Johnson's S_B functions to data from *Quercus robur* stands (N = 172 plots) and *Betula pubescens* (N = 202 plots)

	Parameter *a*	*Quercus robur*				*Betula pubescens*			
		Bias	MAE	MSE	D_n	Bias	MAE	MSE	D_n
Weibull (PW)	d_{min} – 5%	0.001018	0.018537	0.000608	0.1416 (0.0410)	0.001119	0.019446	0.000713	0.1804 (0.0559)
	d_{min} – 10%	0.001047	0.018489	0.000607	0.1396 (0.0399)	0.001161	0.019318	0.000708	0.1787 (0.0553)
	d_{min} – 15%	0.001119	0.018491	0.000609	0.1381 (0.0391)	0.001251	0.019250	0.000707	0.1773 (0.0548)
	d_{min} – 20%	0.001181	0.018489	0.000610	0.1370 (0.0387)	0.001353	0.019211	0.000707	0.1761 (0.0544)
Weibull (ML)	d_{min} – 5%	0.000819	0.018829	0.000622	0.1358 (0.0416)	0.000826	0.020049	0.000761	0.1718 (0.0600)
	d_{min} – 10%	0.000820	0.018758	0.000620	0.1399 (0.0413)	0.000836	0.019823	0.000747	0.1759 (0.0608)
	d_{min} – 15%	0.000870	0.018751	0.000622	0.1427 (0.0414)	0.000912	0.019729	0.000744	0.1788 (0.0612)
	d_{min} – 20%	0.000930	0.018763	0.000625	0.1451 (0.0417)	0.001004	0.019684	0.000745	0.1810 (0.0615)
Weibull (MMW)	d_{min} – 5%	0.000774	0.018812	0.000622	0.1416 (0.0374)	0.000740	0.020077	0.000761	0.1771 (0.0580)
	d_{min} – 10%	0.000849	0.018740	0.000620	0.1414 (0.0365)	0.000847	0.019850	0.000748	0.1772 (0.0579)
	d_{min} – 15%	0.000946	0.018727	0.000622	0.1419 (0.0367)	0.000980	0.019726	0.000742	0.1781 (0.0583)
	d_{min} – 20%	0.001038	0.018731	0.000624	0.1428 (0.0372)	0.001110	0.019659	0.000741	0.1791 (0.0590)
	Parameter ε	Bias	MAE	MSE	D_n	Bias	MAE	MSE	D_n
Johnson's S_B (CML)	d_{min} – 5%	0.000808	0.019515	0.000667	0.1128 (0.0316)	0.000973	0.021845	0.000906	0.1231 (0.0332)
	d_{min} – 10%	0.000667	0.019153	0.000640	0.1139 (0.0332)	0.000761	0.020937	0.000829	0.1260 (0.0381)
	d_{min} – 15%	0.000612	0.018996	0.000629	0.1169 (0.0359)	0.000674	0.020428	0.000791	0.1295 (0.0425)
	d_{min} – 20%	0.000599	0.018916	0.000625	0.1208 (0.0385)	0.000648	0.020117	0.000768	0.1334 (0.0456)
Johnson's S_B (MMJ)	d_{min} – 5%	0.000521	0.019058	0.000637	0.1077 (0.0306)	0.000506	0.021111	0.000845	0.1268 (0.0369)
	d_{min} – 10%	0.000549	0.018878	0.000625	0.1077 (0.0339)	0.000528	0.020585	0.000801	0.1261 (0.0381)
	d_{min} – 15%	0.000598	0.018764	0.000620	0.1096 (0.0374)	0.000576	0.020198	0.000772	0.1272 (0.0407)
	d_{min} – 20%	0.000663	0.018701	0.000617	0.1131 (0.0406)	0.000647	0.019936	0.000754	0.1299 (0.0434)
Johnson's S_B (MJ)	d_{min} – 5%	0.000733	0.024546	0.001798	0.2470 (0.1211)	0.001245	0.033421	0.003122	0.2848 (0.1477)
	d_{min} – 10%	0.000797	0.023624	0.001326	0.2376 (0.1177)	0.001035	0.031955	0.002470	0.2711 (0.1411)
	d_{min} – 15%	0.000926	0.022968	0.001125	0.2267 (0.1119)	0.001114	0.030784	0.002099	0.2583 (0.1309)
	d_{min} – 20%	0.001075	0.022529	0.001015	0.2176 (0.1057)	0.001282	0.029835	0.001862	0.2481 (0.1246)
Jonhson's S_B (KB)	d_{min} – 1.3	0.000391	0.018939	0.000641	0.1295 (0.0460)	0.000441	0.020281	0.000813	0.1594 (0.0506)

d_{min} (cm): minimum observed diameter; PW: percentiles; ML: maximum likelihood; MMW: method of moments for the Weibull function; CML: Conditional Maximum Likelihood; MMJ: method of moments for the Johnson's S_B function; MJ: mode; KB: Knoebel and Burkhart's method.

with the ML method in terms of the KS statistic (D_n), with c = 5% and c = 10% of the minimum observed diameter in the plot. However, in all cases, the best results of MAE and MSE were obtained by the percentiles method (PW), *i.e.* the results were similar to those for pedunculate oak stands. The most appropriate value of the location parameters in terms of MSE were obtained when c = 15% in PW and ML approaches and when c = 20% with MMW. Nevertheless, higher values of the KS statistic were obtained with low values of the location parameter, except in case of the PW, for which the smallest value was obtained with c = 20%.

In case of the Johnson's S_B PDF, the lowest values of D_n were obtained with the Conditional Maximum Likelihood approach (CML), with c = 5% followed by c = 10%. However, the best results in terms of MAE and MSE were obtained with the method of Moments (MMJ) with c = 20%. The method of the Mode (MJ) was not suitable for this stands and in 3 of the plots the mode value was not able to be computed. In relation to the KB, this method was slightly inferior to MMJ and CML in terms of D_n but provided similar values of MAE and MSE as CML and MMJ when c = 10% in both cases.

In both species results in terms of KS statistic were better with the Johnson's S_B function although this four-parameter model is more complex than the three-parameter Weibull distribution.

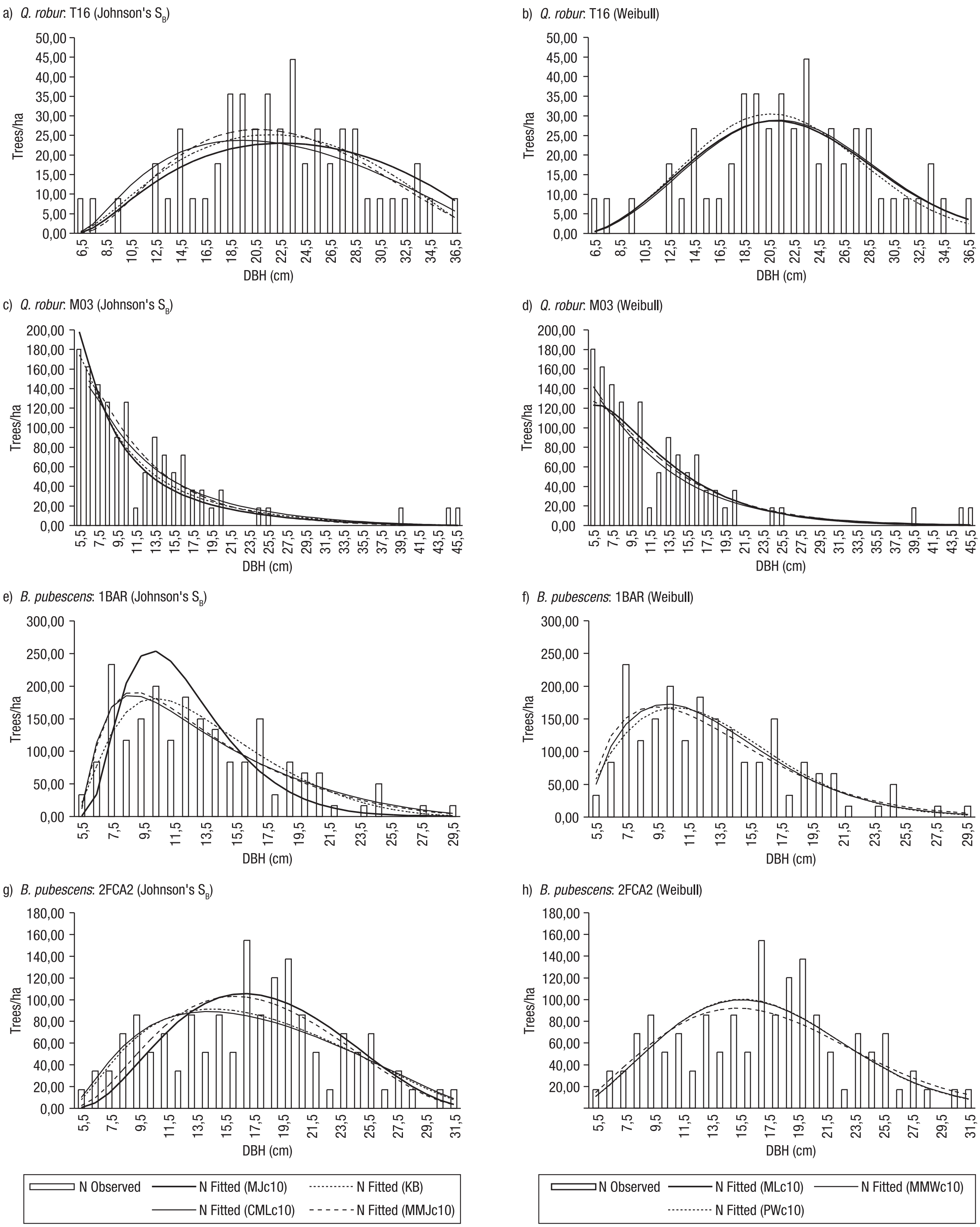

Figure 1 (a, b, c, d, e, f, g and h). Observed and corresponding fitted distributions for 4 plots.

Acknowledgements

The present study was financially supported by the following: Asociación Galega Monte-Industria, within the project "Elaboración de modelos de crecimiento y producción para la gestión sostenible de las masas de abedul" (2000/CG488); the Xunta de Galicia, within the project "Estudio epidométrico de las masas de *Quercus robur* L. en Galicia y su influencia sobre la calidad de la madera" (PGIDT99MA29101); and by the Ministerio de Educación y Ciencia and ERDF, within the project "Modelos dinámicos de crecimiento y capacidad de secuestro de carbono de bosques de frondosas autóctonas del noroeste peninsular" (AGL2007-66739-CO2-01/FOR).

The *Unidade de Xestión Forestal Sostible* (UXFS) is cofunded by the Xunta de Galicia («Programa de Consolidación y Estructuración de Unidades de Investigación Competitivas 2011») and the European Union ERDF programme.

References

Cao Q, 2004. Predicting parameters of a Weibull function for modeling diameter distribution. For Sci 50(5): 682-685.

Diéguez-Aranda U, Grandas-Arias JA, Álvarez-González JG, Gadow Kv, 2006. Site quality curves for birch stands in North-Western Spain. Silva Fennica 40(4): 631-644.

Fonseca TF, Marques CP, Parresol BR, 2009. Describing Maritime Pine Diameter Distributions with Johnson's SB Distribution Using a New All-Parameter Recovery Approach. For Sci 55(4): 367-373.

Gorgoso JJ, Rojo A, Cámara-Obregón A, Diéguez-Aranda U, 2012. A comparison of estimation methods for fitting Weibull, Johnson's S_B and beta functions to Pinus pinaster, Pinus radiata and Pinus sylvestris stands in northwest Spain. Forest Systems 21(3): 446-459.

Gorgoso-Varela JJ, Rojo-Alboreca A, Afif-Khouri E, Barrio-Anta M, 2008. Modelling diameter distributions of birch (*Betula alba* L.) and pedunculate oak (*Quercus robur* L.) stands in northwest Spain with the beta distribution. Invest Agrar: Sist Recur Fo. 17(3): 271-281.

Johnson NL, 1949. Systems of frequency curves generated by methods of translation. Biometrika 36: 149-176.

Knoebel BR, Burkhart HE, 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. Biometrics 47: 241-253.

Palahí M, Pukkala T, Blasco E, Trasobares A, 2007. Comparison of beta, Johnson's S_B, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). Eur J Forest Res 126: 563-571.

Parresol BR, 2003. Recovering parameters of Johnson's S_B distribution. Res. Pap. SRS-31. Asheville, NC: US Department of Agriculture Forest Service, Southern Research Station. 9 pp.

Río M del, 1999. Régimen de claras y modelo de producción para *Pinus sylvestri*s L. en los sistemas Central e Ibérico (dissertation). INIA, Madrid, Spain. 257 pp.

Rojo A, Álvarez-González JG, Grandas JA, 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.

Scolforo JRS, Vitti FC, Grisi RL, Acerbi F, De Assis AL, 2003. S_B distribution's accuracy to represent the diameter distribution of Pinus taeda, through five fitting methods. For Ecol Manage 175: 489-496.

Zhang L, Packard KC, Liu C, 2003. A comparison of estimation methods for fitting Weibull and Johnson's S_B distributions to mixed spruce-fir stands in northeastern North America. Can J For Res 33: 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.