Fitting Johnson's S_B Distribution to Forest Tree Diameter

Ayana Mateus and Margarida Tomé

Abstract

The simulation of diameter distributions is an essential aid for a more efficient planning of the harvesting operations which usually represents a high percentage in costs associated with production of pulp. In this chapter Johnson's S_B probability density function has been used to model diameter distribution of *Eucalyptus globulus* Labill. in Portugal.

1 Introduction

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

The objective of the research report here is to model the diameter distribution of eucalyptus plantations in Portugal.

To achieve this objective the following partial objectives were needed: to identify the probability density function (*pdf*) that better reproduced the set of observed frequencies based on the estimates obtained for coefficients of skewness (β_1) and kurtosis (β_2) in each plot at each measurement age; to develop a system of equations that relates stand basal area (G) with the noncentral moments of the distribution in

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

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

J. Lita da Silva et al. (eds.), *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications*, Studies in Theoretical and Applied Statistics, DOI 10.1007/978-3-642-34904-1_30,

Table 1 Characteristics of tree variable used

Tree variable	Minimum	Mean	Maximum
Diameter at breast height (d, cm)	0.1	11.38	45.40

Table 2 Characteristics of stand variables used

Stand variable	Minimum	Mean	Maximum
Basal area (G, m^2ha^{-1})	0.05	16.16	64.55
Number of trees per hectare (N, ha^{-1})	450	1237.34	2811
Productivity (S, m)	10.33	20.20	33.93
Age (t, years)	0.6	9.36	34.70

order to obtain estimates of the *pdf*. The stand basal area is the sum of squared diameter multiplied by a factor to express it on area per hectare, and it is related to the second noncentral moment of the distribution. This variable expresses the competition between trees as it is the area occupied with tree stems. It has a great importance because in the growth of trees, the competition reflects itself mainly by growth in diameter. The algorithm proposed by Parresol [12] was selected as a starting point for the parameter recovery.

2 Methods

2.1 Data

The data used in this study to model diameter distributions of eucalyptus (*Eucalyptus globulus* Labill.) plantations were collected in Portugal in permanent plots installed in first rotation stands.

The information concerning all the trees within a plot includes successive measurements, usually annually, of diameter at breast height (d).

The plots used in the present research have drawn on a very large data set covering stands with different characteristics, namely age (t), stocking (N, number of trees per hectare), and productivity (S) (see Tables 1 and 2).

2.2 Testing the Performance of Johnson S_B Distribution

The analysis of the coefficients of skewness (β_1) and kurtosis (β_2) of the distribution of the diameters could be used to indicate the appropriate pattern followed by a certain population.

For a first identification of the distribution that better reproduced the set of observed frequencies, the estimates of the coefficients of skewness (β_1) and kurtosis (β_2), in each plot at each measurement age, were first analyzed.

The estimators used were, respectively,

$$b_1 = \frac{m_3}{m_2^{3/2}} \qquad \textit{and} \qquad b_2 = \frac{m_4}{m_2^2}$$

with

$$m_2 = \frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1} \qquad m_j = \frac{\sum_{i=1}^{n}(d_i - \overline{d})^j}{n} \qquad j = 3, 4.$$

d_i is the diameter at breast height of tree i, $\overline{d}$ is the average diameter of the plot, and n is the number of trees measured in the plot.

The choice of the Johnson S_B distribution as the null hypothesis for the modeling of diameter distributions of eucalyptus has been based on its flexibility to model distributions with different shapes. It has a broader range of the (β_1, β_2) space than other distributions and includes most of the alternative *pdf* [5, 6].

Since Hafley and Schreuder [4] introduced the four parameter Johnson's S_B distribution into forest literature, this probability density function has been widely used in forest diameter (and height) distribution modeling by several authors, such as [1, 3, 7, 9, 12, 13, 15, 16].

To test the performance of Johnson S_B distribution to model diameter distributions of eucalyptus plantations in Portugal, the b_1 and b_2 estimates were computed in each plot at each measurement age on the fitting data set in order to check if the pairs (b_1, b_2) occur mainly in the parametric space that corresponds to this distribution [12].

In order to complement the methodology used, based on the analysis of coefficients (β_1, β_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_B distribution fits the diameter distributions on individual plots [11, 14]. We used the modified Kolmogorov–Smirnov test because the parameters were unknown and estimated from the data [10]. The test of the qui-square was not used, for being dependent of the grouping of data in classes.

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 transformation is applied, in order to obtain a normal distribution to the random variable processed. This system is composed by three kinds of distributions (Johnson S_L, S_B, and S_U), depending on the transformation applied to the random variable [5].

When the transformation $Z = \gamma + \delta g(X)$ is made on the random variable X, an infinite system of distribution functions (or random variables) is being defined, clearly identified by the transformation $g(X)$, necessary to obtain a transformed with standard normal distribution.

Johnson introduced four parameters γ, δ, ϵ, and λ, with $\gamma, \epsilon \in \mathbb{R}$, $\lambda \in \mathbb{R}^+$, $\delta \in \mathbb{R} \setminus \{0\}$ and expressed the generic transformation defined above in the following way:

$$Z = \gamma + \delta g\left(\frac{X - \epsilon}{\lambda}\right), \tag{1}$$

where γ and δ are shape parameters and ϵ and λ are location and scale parameters, respectively. Although the parameters γ and δ affecting both the skewness and the kurtosis of distribution, the parameter γ is particularly associated with the asymmetry and an increase in the parameter δ corresponds to an increase in the kurtosis [5].

In order to generate distributions with limited support, the transformed chosen is

$$g(Y) = \ln\left(\frac{Y}{1 - Y}\right) \tag{2}$$

that in terms of the variable $Y = \dfrac{X - \epsilon}{\lambda}$ results in

$$Z = \gamma + \delta \ln\left(\frac{X - \epsilon}{\epsilon + \lambda - X}\right), \quad \epsilon < X < \epsilon + \lambda, -\infty < \gamma < \infty, \delta > 0, -\infty < \epsilon < \infty, \lambda > 0 \tag{3}$$

or

$$Z = \gamma + \delta \ln\left(\frac{Y}{1 - Y}\right), 0 < y < 1, -\infty < \gamma < \infty, \delta > 0, -\infty < \epsilon < \infty, \lambda > 0. \tag{4}$$

The system of random variables generated by Eq. (3) or (4) is called the Johnson S_B system of distributions.

2.4 Algorithm to Estimate the Parameters of the Johnson S_B Distribution

The parameters of the Johnson S_B distribution were estimated using the methodology proposed by Parresol [12].

If Eq. (4) is expressed in terms of the variable Y, the following expression is obtained for Y:

$$Y = \left[1 + \exp\left(-\frac{Z - \gamma}{\delta}\right)\right]^{-1}. \tag{5}$$

When the variable Z assumes the null value the median of the variable Y (or X) is obtained:

$$y_{1/2} = \left(1 + e^{\gamma/\delta}\right)^{-1}. \tag{6}$$

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

Equation (6) enables the estimation of the shape parameter γ, according to the median value of the diameter distribution, provided that the shape parameter δ is known:

$$\gamma = \delta \ln \left(\frac{1}{y_{1/2}} - 1 \right) = \delta \ln \left(\frac{\lambda}{x_{1/2} - \epsilon} - 1 \right). \tag{7}$$

However, another equation is needed to estimate the shape parameter δ. As we said in Sect. 1, a variable of great interest in the elaboration of stand models with diameter distribution simulation is the stand basal area (G). This variable is related to the second noncentral moment, $E(X^2)$ of X, through the relation,

$$G = \frac{1}{10000} \sum_{i=1}^{N} \frac{\pi}{4} d_i^2 = c\ N\ E\left(X^2\right) \qquad (m^2 ha^{-1}) \tag{8}$$

with $N =$ number of trees alive, per hectare, $d_i =$ diameter at breast height (cm) measured of tree i, and $c = \frac{\pi}{40000}$ is a conversion constant.

As

$$E\left(X^2\right) = E\left(\epsilon + \lambda\ Y\right)^2 = \epsilon^2 + 2\epsilon\lambda E\left(Y\right) + \lambda^2 E\left(Y^2\right),$$

then

$$G = c\ N\left(\epsilon^2 + 2\epsilon\lambda E\left(Y\right) + \lambda^2 E\left(Y^2\right)\right). \tag{9}$$

The noncentral moments of order r $(E(Y^r))$ $r = 1, 2$ may be determined through the moment-generating function φ of the variable Y

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

which shows the following relationship:

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

The resolution of the system formed by Eqs. (7) and (9), based on known values of the median variable Y(or X), G, and N, allows, by assuming some reasonable values for ϵ and λ, to obtain estimates for the parameters γ and δ. The solution requires the use of numerical iterative methods of numerical integration, as the calculation of moments of the distribution does not have an analytical solution [2,8].

As in any iterative process it is necessary to assign initial values to the parameters. Parresol [12] suggests the attribution of an initial value to δ for a first approach of γ obtained from Eq. (7). Thus the parameter δ was initialized with the estimate obtained for the kurtosis because an increase of δ corresponds to an increase in the kurtosis [5]. The parameter ϵ was fixed as equal to the minimum value of the observed diameter and λ to the difference between the maximum and minimum value of the observed diameter. The values for G and N were obtained from the measurement of each plot in study.

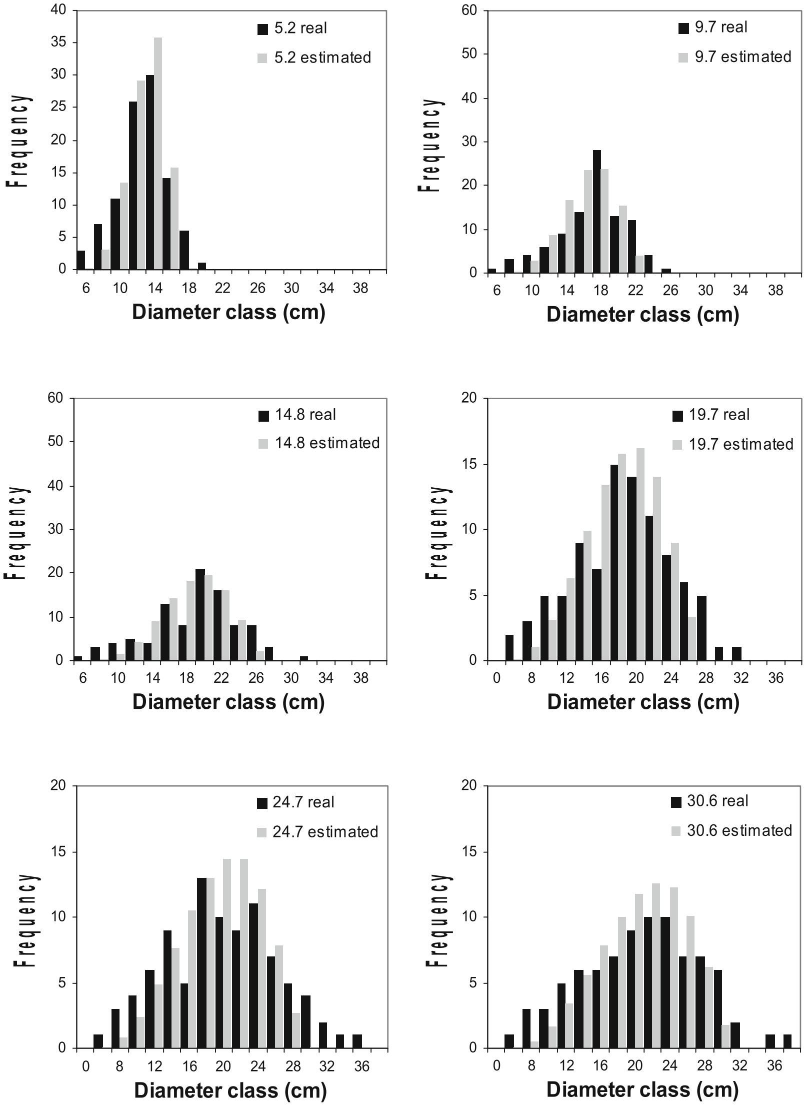

Fig. 1 Comparison of real and estimated diameter distribution from a plot at ages (years) 5.2, 9.7, 14.8, 19.7, 24.7, and 30.6 (*dark* = real values)

3 Results and Conclusions

The ranges for the coefficients (b_1, b_2) estimated with the data set described in Tables 1 and 2 were $-1.3977 \leq b_1 \leq 1.0805$ 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. The values observed for the pairs $(b_1,\ b_2)$ are included in the range of variation for the coefficients of skewness and kurtosis of the Johnson S_B distribution [5].

It was also verified that in the great majority of the plots, the coefficients of skewness assume 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 higher initial growth in diameter (d) will compete, mainly for light, with the smaller ones making those to continue to have lower growth rates, and the differences between small and large trees tended to increase.

The modified Kolmogorov–Smirnov test with a significance level of 5% showed that the distribution Johnson S_B did not significantly differ from the empirical distribution in 106 out of 111 studied stands, each of them with several remeasurements between 5 and 32 years.

In conclusion, modeling diameter distributions of eucalyptus (*Eucalyptus globulus* Labill.) plantations in Portugal through a probability density function, namely Johnson S_B, using a parameter recovery approach seems to be a good methodology that can be generally applied to the most common values of the pair (b_1, b_2). The main advantage of using parameter recovery models is that the stand variable that was used in the parameter recovery, namely basal area, assures compatibility between the characteristics of the observed population and those obtained through simulation of diameter distribution. This means that basal area computed with the simulated distribution is fairly closed to the one observed.

As an example Fig. 1 shows the evolution of the observed and simulated diameter distribution from 5.2 to 30.6 years of age in one of the permanent plot from the fitting data set when the initialization was made with the measurement 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.

Acknowledgements This work was supported by the FP6 EFORWOOD IP project (contract 518128) and the FCT project CarbWoodCork (POCI/AGR/57279/2004). This work was partially supported by CMA/FCT/UNL, under the project PEst-OE/MAT/UI0297/2011. The authors gratefully acknowledge Bernard Parresol (USDA Forest Service Southern Research Station) for facilitating the use of a SAS program for Johnson S_B parameter recovery that was used in an initial stage of this research. The authors also thank Celbi and Silvicaima pulp companies that provided a large proportion of the data used in this study.

References

1. Fonseca, T.F., Marques, C.P., Parresol, B.R.: Describing maritime pine diameter distributions with Johnson's S_B distribution using a new all-parameter recovery approach. For. Sci. **55**(4), 367–373 (2009)
2. Gallant, A.: Nonlinear Statistical Models, p. 624. Wiley, New York (1987)
3. Hafley, W.L., Buford, M.A.: A bivariate model for growth and yield prediction. For. Sci. **31**, 237–247 (1985)

4. Hafley, W.L., Schreuder, H.T.: Statistical distributions for fitting diameter and height data in even-ages stands. Can. J. For. Res. **7**, 481–487 (1977)
5. Johnson, N.L.: Systems of frequency curves generated by methods of translation. Biometrika **36**, 147–176 (1949)
6. Johnson, N., Kotz, S.: Continuous Univariate Distribution, vol. 1, p. 756, 2nd edn. Wiley, New York (1994)
7. Kamziah, A.K., Ahmad, M.I., Lapongan, J.: Nonlinear regression approach to estimating Johnson SB parameters for diameter data. Can. J. For. Res. **29**, 310–314 (1999)
8. Law, A.M., Kelton, W.D.: Simulation Modeling and Analysis. McGraw-Hill, New York (1982)
9. Li, F., Zhang, L., Davis, C.J.: Modelling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. For. Sci. **48**(1), 47–58 (2002)
10. Lilliefors, H.W.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. American Stat. Assoc. J. **62**, 399–402 (1967)
11. Massey, F.J.: The Kolmogorov-Smirnov test for goodness of fit. American Stat. Assoc. J. **46**, 68–78 (1951)
12. Parresol, B.: Recovering parameters of johnson's S_B distribution. Res. pap. SRS-31. Asheville, NC:USDA For. Ser., Southern Research Station, p. 9 (2003)
13. Rennolls, K., Wang, M.: A new parameterization of Johnson's S_B distribution with application to fitting forest tree diameter data. Can. J. For. Res. **35**, 575–579 (2005)
14. Reynolds, M.R., Burk, T.E., Huang, W.: Goodness-of-fit tests and model selection procedures for diameter distributions models. For. Sci. **34**, 377–399 (1988)
15. Zhang, L., Packard, K.C., Liu, C.: A comparison of estimation methods for fitting Weibull and Johnsons SB distribution to mixed spruce-fir stands in northeastern North America. Can. J. For. Res. **33**(7), 1340–1347 (2003)
16. Zhou, B., McTague, J.P.: Comparison and evaluation of five methods of estimation of the Johnson system parameters. Can. J. For. Res. **26**, 928–935 (1996)