iForest
Biogeosciences and Forestry

Research Article
doi: 10.3832/ifor1447-008
vol. 9, pp. 23-29

Modeling extreme values for height distributions in *Pinus pinaster*, *Pinus radiata* and *Eucalyptus globulus* stands in northwestern Spain

J Javier Gorgoso-Varela [1], J Daniel García-Villabrille [2], Alberto Rojo-Alboreca [2]

Methods of estimating extreme height values can be used in forest modeling to improve fits to the marginal distribution of heights in the following bivariate diameter-height models: the S_{BB} Johnson's distribution, the bivariate beta (GDB-2) distribution, the bivariate Logit-Logistic (LL-2) distribution and the power-normal (PN) distribution. Some applications to LiDAR derived data are also possible, *e.g.*, for error calibration. Practical applications in forest management may also be considered, *e.g.*, for pruning. In probability theory and statistics, the generalized extreme value (GEV) distribution, also known as the Fisher-Tippett distribution, is a family of continuous probability distributions that combine the Gumbel, Fréchet and Weibull distributions. This study compared the three distributions for fitting extreme values of tree heights (maximum and minimum heights), which were measured in 185 permanent research plots in *Pinus pinaster* Ait. stands, 97 research plots in *Pinus radiata* D. Don stands, and 128 research plots in *Eucalyptus globulus* Labill. Most of the eucalyptus stands were measured three times giving a total of 304 measurements. All plots are located in northwestern Spain. The Bias, Mean Absolute Error (MAE) and Mean Square Error (MSE) of the mean relative frequency of trees were used to evaluate the goodness-of-fit of the different functions, as well as the Kolmogorov-Smirnov statistic D_n. The Gumbel and the Weibull cumulative distribution functions (CDFs) proved suitable for describing extreme values of height distributions of the above-mentioned tree species in northwestern Spain. The Fréchet distribution was only used to model maximum values and yielded the poorest results in all cases.

Keywords: Gumbel, Fréchet, Weibull, Minimum Height, Maximum Height

Introduction

Methods of estimating extreme values (minimum and maximum tree heights) can be applied in forest modeling, specifically for fitting bivariate height-diameter models such as the bivariate S_{BB} Johnson's distribution (Johnson 1949), the generalized bivariate beta distribution (GDB-2 - Li et al. 2002), the bivariate Logit Logistic distribution (LL-2 - Wang & Rennolls 2007) and the power-normal (PN) distribution (Mønness 2011). The accuracy of these models can be improved during fitting by choosing suitable values of the location and scale parameters, which are related to the minimum and maximum values of the marginal distribution of the S_{BB} (Schreuder & Hafley 1977, Knoebel & Burkhart 1991, Tewari & Von Gadow 1997, 1999, Schmidt & Von Gadow 1999, Zucchini et al. 2001, Li et al. 2002). These values also are used in LiDAR derived information, *e.g.*, to compare the modeled values and the LiDAR-measured heights for error calibration. Knowledge of the extreme values of tree heights in forest stands is also useful for some practical applications such as pruning.

In probability theory and statistics, the generalized extreme value (GEV) distribution, also known as the Fisher-Tippett distribution, is a family of continuous probability distributions that combine the Gumbel, Fréchet and Weibull families of distributions, also known respectively as type I, II and III extreme value distributions (Persson & Rydén 2010). The Gumbel distribution (Gumbel 1954) is used to model the distribution of the maximum and/or the minimum values of a number of samples of various distributions. For example, it could be used to represent the distribution of the maximum level of a river in a particular year when a list of maximum values for the past ten years were available. It is also useful for predicting the probability that an extreme earthquake, flood or other natural disaster will occur. Extreme value theory indicates that the Gumbel distribution is useful for representing the distribution of maximum values when the underlying sample data are normally or exponentially distributed. The Gumbel distribution has been variously called the log-Weibull distribution, the double exponential distribution and the Laplace distribution, and it is often incorrectly referred to as the Gompertz distribution (Willemse & Kaas 2007).

Although the Fréchet distribution is named after Fréchet (1927), who used it to model the distribution of the largest order

(1) Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo, Escuela Politécnica de Mieres, c/ Gonzalo Gutiérrez Quirós s/n, E-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 s/n, E-27002 Lugo, Galicia (Spain)

@ J Javier Gorgoso-Varela (gorgoso@uniovi.es)

Received: Sep 18, 2014 - Accepted: Mar 21, 2015

Citation: Gorgoso-Varela JJ, García-Villabrille JD, Rojo-Alboreca A (2015). Modeling extreme values for height distributions in *Pinus pinaster, Pinus radiata* and *Eucalyptus globulus* stands in northwestern Spain. iForest 9: 23-29. – doi: 10.3832/ifor1447-008 [online 2015-07-25]

Communicated by: Emanuele Lingua

statistic, it was further developed by Fisher & Tippett (1928) and Gumbel (1954). It has been shown to be useful in accelerated life testing and for modeling and analyzing rainfall, sea currents and wind speeds, as well as extreme events such as earthquakes and floods. In hydrology, the Fréchet distribution is used to model extreme events such as annual maximum one-day rainfall and river discharges (Coles 2001). Finally, the Weibull distribution (Weibull 1951) is a simple, flexible model and it is often used in forestry studies involving diameter distributions (Bailey & Dell 1973, Rennolls et al. 1985, Maltamo et al. 1995, Nanos & Montero 2002, Zhang et al. 2003, Liu et al. 2004, Palahí et al. 2007). It is also used to model extreme values in many scientific disciplines.

The objective of the present study was to fit the three extreme value distributions (Gumbel, Fréchet and Weibull) to maximum and minimum tree heights in *P. radiata, P. pinaster* and *E. globulus* stands in northwestern Spain. The distributions were fitted separately as independent functions, and not jointly to yield the Generalized Extreme Value distribution.

Material and methods

Data set

Maritime pine (*P. pinaster* Ait.), Monterrey pine (*P. radiata* D. Don) and blue gum (*E. globulus* Labill.) stands represent three of the most important forest resources in northwestern Spain. These species mainly occur in pure stands, although they sometimes may be found in mixed stands, and are the most commonly used species in productive stands in this area. Pure stands of maritime pine cover 217 281 ha in the region of Galicia and 22 523 ha in the adjoining region of Asturias. These stands are mainly derived from natural regeneration and occasionally from plantations. Exotic Monterrey pine plantations cover 96 177 ha in Galicia and 25 385 ha in Asturias (MMAMRM 2011). Pure *E. globulus* stands cover an area of 320 774.81 ha, 100 245.72 ha as mixed *E. globulus* and *P. pinaster* Ait. stands and 12 895.30 ha as mixed *E. globulus* and *Q. robur* L. stands in Galicia. Pure stands of *E. globulus* cover an area of 60 000 ha in Asturias (MMAMRM 2011).

The data used in this study correspond to 185 permanent research plots established in maritime pine (*P. pinaster*) stands, 97 plots in Monterrey pine (*P. radiata*) stands and 128 research plots in blue gum (*E. globulus*) stands. Most eucalyptus stands were measured three times, giving a total of 304 measurements. Due to the fast growth of *E. globulus*, the measurements are not considered repetitive and re-measurement does not affect the results. All plots are located in stands in NW Spain (in the regions of Galicia and Asturias), except for some *P. radiata* plots that are located in a small area of Castilla y León. In the *P. pinaster* stands, the plot size ranged from 375 to 900 m^2, depending on the stand density, with a minimum of 30 trees per plot. In the *P. radiata* stands the plot size was 1000 m^2, while in the *E. globulus* stands the plot size was about 500 m^2. The minimum and the maximum heights in each distribution were extracted to form the experimental distributions of extreme values for each species. These distributions were then used for model parametrization.

The research plots used in the present study were established in stands dominated by the subject tree species (more than 85% of standing basal area) in order to cover a wide variety of combinations of age, number of trees per hectare and sites. All trees in each plot were numbered. The heights were measured with a Vertex IV hypsometer to the nearest 0.1 m. The empirical data represent non-truncated distributions.

The following stand variables were calculated from the inventory data: quadratic mean diameter, number of trees per ha, basal area and dominant height. Summary statistics of the main stand variables are shown in Tab. 1.

Model fitting

The Generalized Extreme Value (GEV) distribution (Fisher & Tippett 1928) has the following cumulative distribution function (CDF) for a random variable x (eqn. 1):

$$F(x)=\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

for $1+\xi[(x-\mu)/\sigma]>0$, where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. The shape parameter (ξ) governs the tail behavior of the distribution. The sub-families defined by $\xi=0$, $\xi>0$ and $\xi<0$ correspond, respectively, to the Gumbel, Fréchet and Weibull families, although the reversed Weibull is the model used to combine the three distributions in the GEV. In the present study, the distributions were fitted independently (and not jointly to yield the GEV).

The Gumbel distribution

The probably density function (PDF) and the cumulative distribution function (CDF - Gumbel 1954) are formulated for a random variable as follows (eqn. 2, eqn. 3):

$$PDF: f(x)=\frac{1}{\beta}\exp[-(z+\exp(-z))]$$

$$CDF: F(x)=\exp\left[-\exp\left(-\frac{x-\mu}{\beta}\right)\right]$$

where $z=(x-\mu)/\beta$, $-\infty<x<\infty$, μ is the mode value (location parameter), β is the scale parameter, and the standard deviation (σ) is (eqn. 4):

$$\sigma=\frac{\beta\pi}{\sqrt{6}}$$

The function was fitted using a location parameter ($\hat{\mu}$) recovered from the experimental mean ($\bar{d}$) and standard deviation (σ) of the distributions, with the following expression (eqn. 5):

$$\hat{d}=\hat{\mu}+\beta\gamma$$

where γ is the Euler-Mascheroni constant (eqn. 6):

$$\gamma=\lim_{n\to\infty}\left[\sum_{k=1}^{n}\frac{1}{k}-\ln(n)\right]=\int_{1}^{\infty}\left(\frac{1}{\lfloor x\rfloor}-\frac{1}{x}\right)dx\approx 0.577215$$

The Fréchet distribution

The Fréchet distribution is a special case of the generalized extreme value distribution. This type-II extreme value distribution is equivalent to the inverse values of a standard Weibull distribution. The probability density function (PDF) and the cumulative distribution function (CDF) for the Fréchet distribution (Fréchet 1927) used for the largest order statistic are as follows (eqn. 7, eqn. 8):

$$PDF: f(x)=\frac{\alpha}{s}\left(\frac{x-m}{s}\right)^{-1-\alpha}\exp\left[-\left(\frac{x-m}{s}\right)^{-\alpha}\right]$$

$$CDF: F(x)=\exp\left[-\left(\frac{x-m}{s}\right)^{-\alpha}\right]\quad \text{if } x>m$$

where $\alpha>0$ is the shape parameter. In this case, the distribution is generalized to in-

Tab. 1 - Descriptive statistics of main variables for the stands analyzed. (*dg*): quadratic mean diameter (cm); (*N*): trees ha^{-1}; (H_0): dominant height (m); (*G*:): basal area (m^2 ha^{-1}); (SD): standard deviation.

Species	Variable	Mean	Max	Min	SD
Pinus pinaster (N=185)	*dg*	20.3	41.5	10.4	7.1
	N	1245.1	2480.0	375.0	483.4
	H_0	14.1	30.6	5.4	4.7
	G	36.1	76.25	7.8	14.9
Pinus radiata (N=97)	*dg*	17.8	25.6	13.4	3.1
	N	1253.6	2543.3	596.7	357.0
	H_0	18.2	27.0	13.3	3.1
	G	30.0	44.3	17.6	6.6
Eucalyptus globulus (N=304)	*dg*	13.4	34.8	1.5	4.3
	N	1174.4	2386.8	435.7	339.1
	H_0	19.1	40.1	3.1	6.3
	G	16.5	63.5	0.15	8.3

clude a location parameter m (the minimum value of the distribution) and a scale parameter $s > 0$.

Parameters of the Fréchet distribution were obtained using the method of moments: mean $\bar{d}$ and variance σ^2, with the following equations (eqn. 9):

$$\bar{d} = m + s\,\Gamma\left(1 - \frac{1}{\alpha}\right)$$

for α>1 and (eqn. 10):

$$\sigma^2 = s^2\left\{\Gamma\left(1 - \frac{2}{\alpha}\right) - \left[\Gamma\left(1 - \frac{1}{\alpha}\right)\right]^2\right\}$$

for α>2. Eqn. 9 and 10 were solved using iterative procedures with the solver function of Microsoft Excel®.

The Weibull distribution

The CDF for the three-parameter Weibull CDF is obtained by integrating the Weibull PDF. For a continuous random variable x it has the following expression (eqn. 11, eqn. 12):

$$PDF: f(x, \mu, \beta, \alpha) = \\ = \frac{\alpha}{\beta}\left(\frac{x-\mu}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]$$

$$CDF: F(x; \mu, \beta, \alpha) = 1 - \exp\left[-\left(\frac{x-\mu}{\beta}\right)^{\alpha}\right]$$

where μ is the location parameter, β is the scale parameter and α is the shape parameter. The scale parameter β and the shape parameter α of the Weibull distribution were obtained by the method of moments. The location parameter μ was predetermined as the minimum value in each distribution, and 1 m height classes were used in the distributions.

In this study, the method of moments was chosen for fitting because the moments of the distribution were also used for fitting the Gumbel and Fréchet CDFs. Such method has been previously applied (Nanang 1998, Del Río 1999, Stankova & Zlatanov 2010, Gorgoso et al. 2012), and is based on the relationship between the parameters of the Weibull function and the first and second moments of the diameter distribution (mean diameter and variance, respectively – eqn. 13, eqn. 14):

$$\beta = \frac{\bar{d} - \mu}{\Gamma\left(1 + \frac{1}{\alpha}\right)}$$

$$\sigma^2 = \frac{(\bar{d} - \mu)^2}{\Gamma^2\left(1 + \frac{1}{\alpha}\right)} \cdot \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$$

where $\bar{d}$ is the arithmetic mean diameter of the distribution, σ^2 is the variance and Γ is the Gamma function. Eqn. 14 was resolved by a bisection iterative procedure (Gerald & Wheatley 1989).

All distributions were fitted using the software SAS/STAT™ (SAS Institute Inc 2003).

Goodness-of-fit evaluation

The consistency of the model and the fitting method were evaluated using the Kolmogorov-Smirnov statistic (D_n). For a given cumulative distribution function $F(x)$: $D_n = sup_x|F_n(x) - F_o(x)|$, where sup_x is the supreme of the set of distances. This value was calculated as follows (eqn. 15):

$$D_n = max\left\{\begin{matrix} max_{1\leq i\leq n_i} F_n(x_i) - F_0(x_j), \\ max_{1\leq i\leq n_i} F_0(x_j) - F_n(x_{i-1}) \end{matrix}\right\}$$

where the cumulative observed frequency $F_n(x_i)$ is compared with the cumulative estimated frequency $F_o(x_j)$.

Bias, mean absolute error (MAE) and mean square error (MSE) were also used as goodness-of-fit measures and were expressed as follows (eqn. 16, eqn. 17, eqn. 18):

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

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

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

where Y_i is the relative frequency of trees observed value in each height 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 as mean relative frequency of trees.

Tab. 2 - Main descriptive statistics of the distributions studied: mean, maximum and minimum values, 25th, 50th and 75th percentiles, standard deviation and skewness and kurtosis coefficients. (h_{max}); maximum height; (h_{min}): minimum height; (P_{25}): 25th percentile; (P_{50}): 50th percentile; (P_{75}): 75th percentile; (Sk): skewness coefficient; (Kur): kurtosis coefficient.

Species	Variable	Mean	Max	Min	P_{25}	P_{50}	P_{75}	SD	Sk	Kur
Pinus pinaster	h_{max}	16.32	36.0	6.1	12.2	15.5	18.9	5.59	0.95	0.81
	h_{min}	7.41	16.7	1.4	4.9	6.6	9.5	3.23	0.72	-0.14
Pinus radiata	h_{max}	21.13	31.7	15.2	18.4	20.7	23.2	3.45	0.72	0.56
	h_{min}	6.40	13.6	2.6	4.5	6.2	7.5	2.22	0.81	0.69
Eucalyptus globulus	h_{max}	20.98	42.6	3.3	16.3	20.9	25.2	6.91	0.32	0.49
	h_{min}	5.23	14.3	0.6	3.5	4.8	6.9	2.48	0.65	0.41

Results

The main descriptive statistics of the distributions under study, including the mean, maximum and minimum values, 25th, 50th and 75th percentiles, standard deviation, and skewness and kurtosis coefficients, are summarized in Tab. 2. The parameter estimates of the Fréchet, Gumbel and Weibull distributions are shown in Tab. 3. The mean values of bias, mean absolute error (MAE) and mean square error (MSE) in relative frequency of trees and the Kolmogorov-Smirnov statistic (D_n) for the fits with the distributions in forest stands of the three species in NW Spain are shown in Tab. 4.

The mean value of the mean square error (MSE) in the relative frequency of number of trees in each diameter class for the fits with the three distributions is shown in Fig. 1. The observed and predicted distributions of maximum and minimum heights for maritime pine, Monterrey pine and blue gum are shown in Fig. 2.

Results showed that the Gumbel and the Weibull CDFs are suitable for describing extreme tree heights in *P. pinaster, P. radiata* and *E. globulus* stands in northwestern

Tab. 3 - Parameter values for the Gumbel, Fréchet and Weibull distributions fitted using the moments approach. (h_{max}); maximum height; (h_{min}): minimum height; (a): values of the parameter m; (b): values of the parameter s.

Distribution	Species	Variable	μ	β	α
Gumbel	*Pinus pinaster*	h_{max}	13.81	4.36	-
		h_{min}	5.96	2.52	-
	Pinus radiata	h_{max}	19.57	2.69	-
		h_{min}	5.40	1.73	-
	Eucalyptus globulus	h_{max}	17.88	5.39	-
		h_{min}	4.12	1.93	-
Fréchet	*Pinus pinaster*	h_{max}	6.1 [a]	7.93 [b]	3.39
	Pinus radiata	h_{max}	15.2 [a]	4.62 [b]	3.30
	Eucalyptus globulus	h_{max}	3.3 [a]	14.66 [b]	4.24
Weibull	*Pinus pinaster*	h_{max}	6.1	11.52	1.90
		h_{min}	1.4	6.78	1.94
	Pinus radiata	h_{max}	15.2	6.66	1.77
		h_{min}	2.6	4.27	1.76
	Eucalyptus globulus	h_{max}	3.3	19.87	2.77
		h_{min}	0.6	5.23	1.95

Tab. 4 - Mean values of bias, mean absolute error (MAE), mean square error (MSE) and Kolmogorov-Smirnov statistic (D_n). (h_{max}); maximum height; (h_{min}): minimum height.

Species	Variable	Distribution	Bias	MAE	MSE	D_n
Pinus pinaster	h_{max}	Gumbel	0.01353	0.01662	0.00053	0.06139
		Fréchet	0.01006	0.04593	0.00478	0.20288
		Weibull	0.01406	0.01863	0.00076	0.07219
	h_{min}	Gumbel	0.02630	0.02931	0.00182	0.09556
		Weibull	0.02767	0.02873	0.00176	0.09507
Pinus radiata	h_{max}	Gumbel	0.02745	0.02896	0.00181	0.11332
		Fréchet	0.02871	0.06182	0.00991	0.27513
		Weibull	0.02838	0.02935	0.00147	0.09595
	h_{min}	Gumbel	0.03655	0.03655	0.00211	0.12012
		Weibull	0.03718	0.03718	0.00209	0.11979
Eucalyptus globulus	h_{max}	Gumbel	0.01033	0.02316	0.00089	0.12639
		Fréchet	0.00713	0.04421	0.00414	0.18785
		Weibull	0.01154	0.01666	0.00050	0.07409
	h_{min}	Gumbel	0.02965	0.02965	0.00161	0.12046
		Weibull	0.03003	0.03007	0.00165	0.11230

Spain (Tab. 4, Fig. 1, Fig. 2). The Fréchet distribution used for the maximum values yielded the poorest results in all cases under study. It tended to underestimate frequencies in the lower half of the data range and then to overestimate the frequencies in the upper half.

The Gumbel and the Weibull distributions generally yielded similar results for the fits to distributions of minimum heights, as indicated by the main statistics used to compare the results (Kolomogorov-Smirnov D_n statistic and mean square error, MSE). The bias may be less useful because errors with different signs tend to cancel

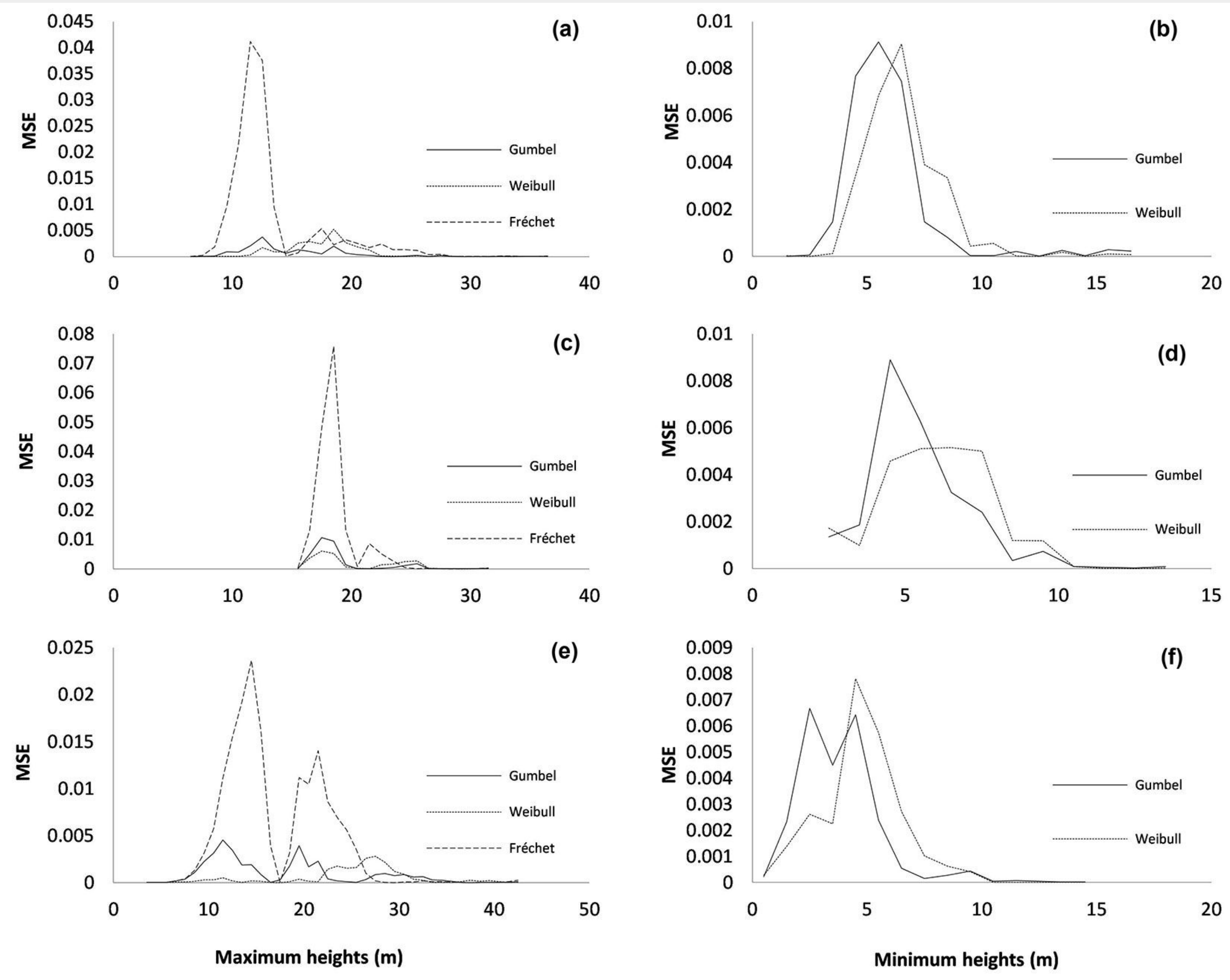

Fig. 1 - Mean values of MSE in each height class in the fits with the Gumbel, Weibull and Fréchet CDFs used for describing extreme values of heights in three tree species from NW Spain. (a-b): *Pinus pinaster*; (c-d): *Pinus radiata*; (e-f): *Eucalyptus globulus*; (a-c-e): maximum heights (m); (b-d-f): minimum heights (m).

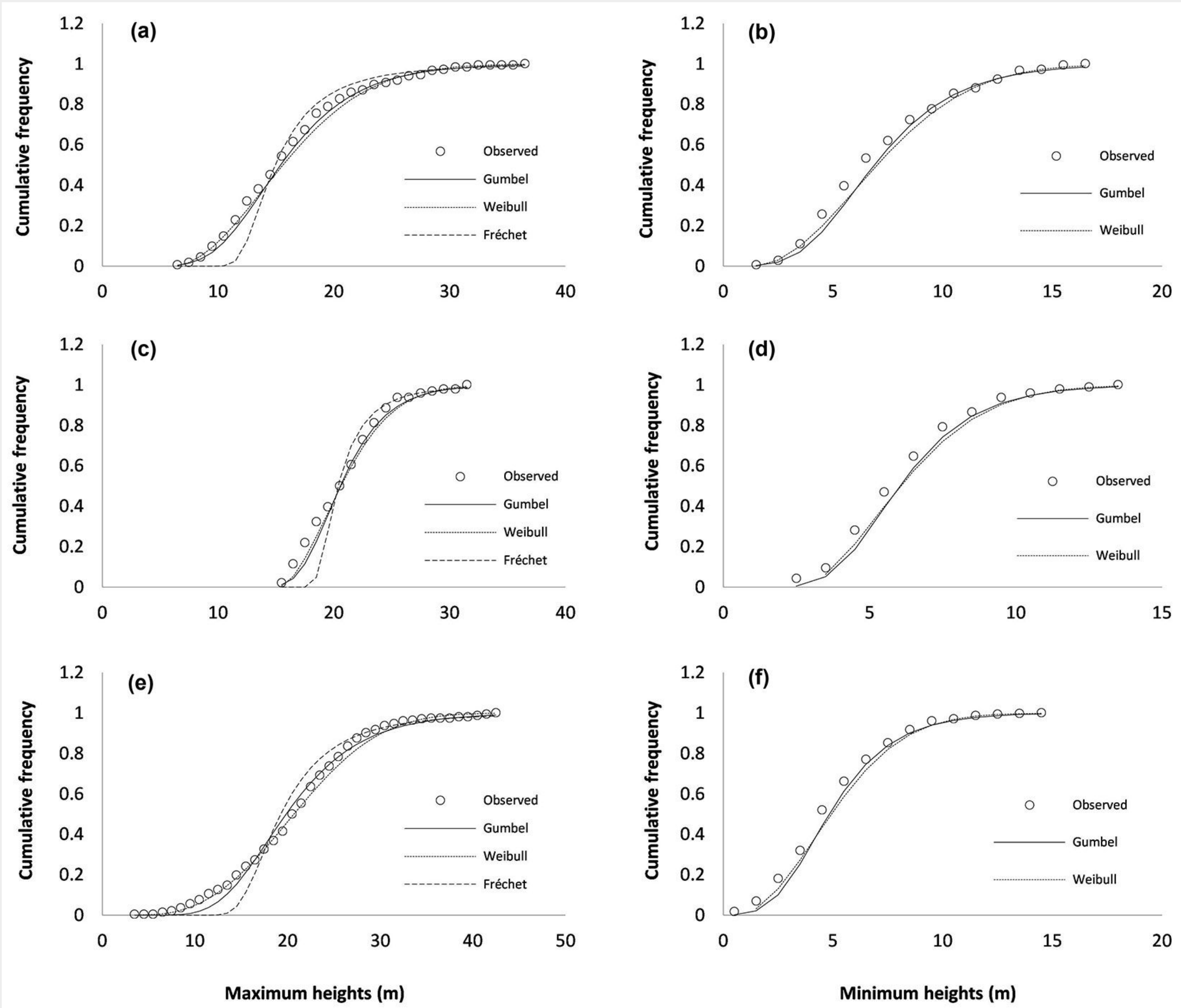

Fig. 2 - Observed and fitted cumulative distributions, for which the Gumbel, Weibull and Fréchet CDFs were used to describe extreme values of heights in three tree species from NW Spain. (a-b): *Pinus pinaster*; (c-d): *Pinus radiata*; (e-f): *Eucalyptus globulus*; (a-c-e): maximum heights (m); (b-d-f): minimum heights (m).

out, thus confounding the overall value. The results were slighter better for maximum than for minimum heights. The Gumbel CDF was the most suitable model for *P. pinaster* stands, while the Weibull CDF was the most appropriate for *P. radiata* and *E. globulus* stands.

Discussion

In this study, extreme (maximum and minimum) tree height data from permanent plots of *P. pinaster, P. radiata* and *E. globulus* species – representing the heterogeneity and complexity of forest stands in the study area – were fitted using three extreme value distributions. This is a novel approach in forest modeling.

Knowledge of the distributions of the maximum and minimum tree heights in forest stands is useful in forest modeling, for example, for improving fits of the bivariate distribution functions. In the Johnson's S_{BB} distribution, the location parameter (ε_h) of the Johnson's S_B marginal distribution of heights is usually fixed as the minimum height of the distribution, while the scale parameter (λ_h) of the same marginal distribution of heights is considered as the range of the distributions, *i.e.*, as maximum height - minimum height (Schreuder & Hafley 1977, Tewari & Von Gadow 1999, Li et al. 2002, Zucchini et al. 2001). Some authors have considered a value of 1.3 for the location parameter when fitting the marginal distribution of heights (Siipilehto 2000, Castedo-Dorado et al. 2001). However, Mønness (2011) compared the Johnson's S_B and the power-normal (PN) distributions fitted to diameters and heights of trees in forest stands by fixing the location parameter of the Johnson's S_B distribution of heights as $H_{min} - (H_{max} - H_{min})/n$.

The final accuracy of the bivariate S_{BB} distribution could be improved by increasing the accuracy of the fits of the Johnson's S_B marginal distributions of diameters and heights. As for diameters, several studies have fixed different location parameters of the Johnson's S_B in relation to the minimum diameter of the distribution (Knoebel & Burkhart 1991, Zhang et al. 2003, Parresol 2003, Fonseca et al. 2009, Gorgoso et al. 2012). However, similar studies have not been applied to the S_{BB} model, which could be improved by a similar approach applied to the marginal distribution of heights. In this case, knowledge of the distribution of minimum heights may be useful for choosing the most suitable location parameter (ε_h) instead of trying to use complicated algorithms to predetermine it. The value of the scale parameter of the extreme value distributions fitted in the present study could help in choosing the ideal location parameter in these marginal distributions by fixing it as a fraction of the minimum diameter observed: this fraction is small when the scale parameter of the extreme value distribution is low (Gorgoso-Varela &

Rojo-Alboreca 2014).

Similar applications of extreme height distributions could also be used with the generalized beta distribution (GDB-2) and the marginal distribution of both diameters and heights, as reported by Li et al. (2002). These authors estimated two of the four parameters of each marginal GDB-2 (β_1, β_2, β_3 and β_4) by substituting $\hat{\beta}_1 = x_{min}$ and $\hat{\beta}_2 = x_{max} - \hat{\beta}_1$, where x_{min} and x_{max} are the minimum and the maximum values of the marginal distributions of heights and diameters. Wang & Rennolls (2007) presented a bivariate distribution (LL-2) based on the univariate Logit-Logistic (Wang & Rennolls 2005), which is obtained by the parametrization of the Johnson's S_B distribution. The location parameter is the same in both distributions.

Airborne light detection and ranging (LiDAR) also uses maximum and minimum heights and has proven useful for characterizing the forest canopy in three dimensions (Watt et al. 2013). Since the first application of airborne LiDAR in forestry over a decade ago (Nilsson 1996), the technology has been widely used to quantify the spatial variation in tree height and crown dimensions at resolutions ranging from stand level (Hall et al. 2005, Naesset & Bjerknes 2001), to plot level (Holmgren et al. 2003, Lim & Treitz 2004, Popescu et al. 2004) and individual tree level (Chen et al. 2006, Coops et al. 2004, Holmgren & Persson 2004, Popescu & Zhao 2008, Roberts et al. 2005). Comparison of the maximum heights estimated from the extreme value distributions with the maximum heights measured by LiDAR at individual tree level is useful, mainly for error calibration, which enables recalculation of all heights measured by LiDAR and estimation of the stand structure.

The following procedure is commonly used to correct errors in LiDAR derived data. The LiDAR data are obtained for the study area and the tree heights are measured in the field (usually with a Vertex hypsometer). The LiDAR technique is used to construct a Digital Terrain Model (DTM) and a Digital Surface Model (DSM) and to determine the tree heights from the vertical difference between such models (DSM-DTM). The LiDAR-derived heights are usually smaller than the field-measured heights. A regression model is then fitted to both sets of height data to correct the LiDAR-measured heights. The accuracy of this model is assessed using the coefficient of determination.

Several other applications of our results may be also identified, such as pruning. The most appropriate timing of pruning is a very important decision that depends on the height and/or diameter of the tree. In general, the height criterion is easier and cheaper to establish in the field. Pruning is carried out in *P. pinaster* and *P. radiata* stands in NW Spain to improve the quality of the wood mainly for the saw and veneer industries. To prevent growth reduction, less than 33% of the total height of the tree should be removed in young stands and less than 50% in old stands. Thus, the first pruning may be applied to trees of minimum height 8 m, with 2.5-2.7 m of the tree removed.

Conclusion

In conclusion, the three extreme value distributions (Gumbel, Fréchet and Weibull) were fitted independently to observed maximum and minimum heights in *P. pinaster, P. radiata* and *E. globulus* stands in northwestern Spain, in a novel approach in the field of forest modeling. The most common potential applications are in forest modeling, for fitting bivariate height-diameter distributions, and in other fields where extreme values are used, such as LiDAR. Practical applications in forest management may also be considered, for example, for pruning.

Acknowledgements

The authors thank forestry students from the Universities of Oviedo and Santiago de Compostela who participated in the fieldwork.

The study was financially supported by the *Gobierno del Principado de Asturias* through the project entitled "*Estudio del crecimiento y producción de Pinus pinaster Ait. en Asturias*" (CN-07-094); by the *Ministerio de Ciencia e Innovación* through the project entitled "*Influencia de los tratamientos selvícolas de claras en la producción, estabilidad mecánica y riesgo de incendios forestales en masas de Pinus radiata D. Don y Pinus pinaster Ait. en el Noroeste de España*" (AGL2008-02259), and an ongoing research project entitled "Growth and yield modelling of clonal and seedling plantations of *Eucalyptus globulus* Labill. of NW Spain" (code AGL2010-22308-C02-01), funded by the Ministry of Science and Innovation of Spain and the European Union through the ERDF programme for the period 2011-2013.

The Sustainable Forest Management Unit (UXFS) is funded by the Xunta de Galicia ("Consolidation and Structuring Program of Competitive Research Units 2011") and by the ERDF.

References

Bailey RL, Dell TR (1973). Quantifying diameter distributions with the Weibull function. Forest Science 19: 97-104. [online] URL: http://www.ingentaconnect.com/content/saf/fs/1973/00000019/00000002/art00004

Castedo-Dorado F, Ruiz-Gonzalez AD, Alvarez-González JG (2001). Modelización de la relación altura-diámetro para *Pinus pinaster* Ait. en Galicia mediante la función de densidad bivariante SBB [Modeling the height-diameter relationship for *Pinus pinaster* Ait. in Galicia using the bivariate SBB function]. Investigación Agraria: Sistemas y Recursos Forestales 10 (1): 111-125. [in Spanish] [online] URL: http://recyt.fecyt.es/index.php/IA/article/viewArticle/2582

Chen Q, Baldocchi D, Gong P, Kelly M (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. Photogrammetric Engineering and Remote Sensing 72: 923-932. - doi: 10.14358/PERS.72.8.923

Coles S (2001). An introduction to statistical modeling of extreme values. Springer-Verlag, London, UK, pp. 209. - doi: 10.1007/978-1-4471-3675-0

Coops NC, Wulder MA, Culvenor DS, St-Onge B (2004). Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data. Canadian Journal of Remote Sensing 30: 855-866. - doi: 10.5589/m04-045

Del Río M (1999). Régimen de claras y modelo de producción para *Pinus sylvestris* L. en los sistemas Central e Ibérico [A thinning program and yield model for *Pinus sylvestris* L. in Spanish Central and Iberian Ranges]. PhD Thesis, Serie Forestal 2, INIA, Madrid, Spain, pp. 257. [in Spanish]

Fisher RA, Tippett LHC (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 24: 190-190. - doi: 10.1017/S0305004100015681

Fonseca TF, Marques CP, Parresol BR (2009). Describing maritime pine diameter distributions with Johnson's SB distribution using a new all-parameter recovery approach. Forest Science 55(4): 367-373. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2009/00000055/00000004/art00008

Fréchet M (1927). Sur la loi de probabilité de l'écart maximum [On the probabilistic law of maximum deviance]. Annales de la Sociéte Polonaise de Mathematique 6: 93. [in French]

Gerald CF, Wheatley PO (1989). Applied numerical analysis (4th edn). Addison-Wesley Publishing Co, Reading, MS, USA, pp. 597.

Gorgoso JJ, Rojo A, Cámara-Obregón A, Diéguez-Aranda U (2012). A comparison of estimation methods for fitting Weibull, Johnson's SB and beta functions to *Pinus pinaster*, *Pinus radiata* and *Pinus sylvestris* stands in northwest Spain. Forest Systems 21 (3): 446-459. - doi: 10.5424/fs/2012213-02736

Gorgoso-Varela JJ, Rojo-Alboreca A (2014). Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands. Annals of Forest Science 71: 741-750. - doi: 10.1007/s13595-014-0369-1

Gumbel EJ (1954). Statistical theory of extreme values and some practical applications. Applied Mathematics Series 33, US Department of Commerce, National Bureau of Standards, Washington, DC, USA, pp. 51.

Hall SA, Burke IC, Box DO, Kaufmann MR, Stoker JM (2005). Estimating stand structure using discrete-return LiDAR: an example from low density, fire prone ponderosa pine forests. Forest Ecology and Management 208 (1-3): 189-209. - doi: 10.1016/j.foreco.2004.12.001

Holmgren J, Persson A (2004). Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment 90 (4): 415-423. - doi: 10.1016/S0034-4257(03)00140-8

Holmgren J, Nilsson M, Olsson H (2003). Estimation of tree height and stem volume on plots using airborne laser scanning. Forest Science 49: 419-428. [online] URL: http://www.ingenta

connect.com/content/saf/fs/2003/00000049/00000003/art00009
Johnson NL (1949). Bivariate distributions based on simple translation systems. Biometrika 36: 297-304. - doi: 10.1093/biomet/36.3-4.297
Knoebel BR, Burkhart HE (1991). A bivariate distribution approach to modeling forest diameter distributions at two points in time. Biometrics 47: 241-253. - doi: 10.2307/2532509
Li F, Zhang L, Davis CJ (2002). Modeling the joint distribution of tree diameters and heights by bivariate generalized Beta distribution. Forest Science 48 (1): 47-58. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2002/00000048/00000001/art00005
Lim KS, Treitz PM (2004). Estimation of aboveground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research 19: 558-570. - doi: 10.1080/02827580410019490
Liu C, Zhang SY, Lei Y, Newton PF, Zhang L (2004). Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. Canadian Journal of Forest Research 34: 2424-2432. - doi: 10.1139/x04-117
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*. Scandinavian Journal of Forest Reseach 10: 284-295. - doi: 10.1080/02827589509382895
MMAMRM (2011). Cuarto Inventario Forestal Nacional [Fourth National Forest Inventory]. Ministerio de Medio Ambiente y Medio Rural y Marino, Galicia, Spain, pp. 52. [in Spanish]
Mønness E (2011). The power-normal distribution: application to forest stands. Canadian Journal of Forest Research 41: 707-714. - doi: 10.1139/x10-246
Naesset E, Bjerknes KO (2001). Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sensing of Environment 78 (3): 328-340. - doi: 10.1016/S0034-4257(01)00228-0
Nanang DM (1998). Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of neem plantations in Northern Ghana. Forest Ecology and Management 103: 1-7. - doi: 10.1016/S0378-1127(97)00155-2
Nanos N, Montero G (2002). Spatial prediction of diameter distributions models. Forest Ecology and Management 161: 147-158. - doi: 10.1016/S0378-1127(01)00498-4
Nilsson M (1996). Estimation of tree heights and stand volume using an airborne LiDAR system. Remote Sensing of Environment 56 (1): 1-7. - doi: 10.1016/0034-4257(95)00224-3
Palahí M, Pukkala T, Blasco E, Trasobares A (2007). Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). European Journal of Forest Research 126: 563-571. - doi: 10.1007/s10342-007-0177-3
Parresol BR (2003). Recovering parameters of Johnson's SB distribution. Research Paper SRS-31, Southern Research Station, USDA Forest Service, Ashville, NC, USA, pp. 9. [online] URL: http://www.treesearch.fs.fed.us/pubs/5455
Persson K, Rydén J (2010). Exponentiated Gumbel distribution for estimation of return levels of significant wave height. Journal of Environmental Statistics 1 (3): 1-12. [online] URL: http://www2.math.uu.se/research/pub/Ryden2.pdf
Popescu SC, Zhao K (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment 112: 767-781. - doi: 10.1016/j.rse.2007.06.011
Popescu SC, Wynne RH, Scrivani JA (2004). Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. Forest Science 50: 551-565. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2004/00000050/00000004/art00013
Rennolls K, Geary DN, Rollinson TJ (1985). Characterizing diameter distributions by the use of the Weibull distribution. Forestry 58: 57-66. - doi: 10.1093/forestry/58.1.57
Roberts SD, Dean TJ, Evans DL, McCombs JW, Harrington RL, Glass PA (2005). Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derived measurements of height and crown dimensions. Forest Ecology and Management 213 (1-3): 54-70. - doi: 10.1016/j.foreco.2005.03.025
SAS Institute Inc (2003). SAS/STAT™ user's guide (Version 9.1). Cary, NS, USA, pp. 409.
Schmidt VM, Von Gadow K (1999). Baumhöhenschätzung mit Hilfe der bivariaten Johnson's SBB-Funktion [Individual tree high estimation by using the bivariate Johnson's SBB function]. Forstw. Cbl. 118: 355-367. [in German] - doi: 10.1007/BF02768999
Schreuder HT, Hafley WL (1977). A useful bivariate distribution for describing stand structure of tree heights and diameters. Biometrics 33: 471-478. - doi: 10.2307/2529361
Siipilehto J (2000). A comparison of two parameter prediction methods for stand structure in Finland. Silva Fennica 34 (4): 331-349. - doi: 10.14214/sf.617
Stankova TV, Zlatanov TM (2010). Modeling diameter distribution of Austrian black pine (*Pinus nigra* Arn.) plantations: a comparison of the Weibull frequency distribution function and percentile-based projection methods. European Journal of Forest Research 129: 1169-1179. - doi: 10.1007/s10342-010-0407-y
Tewari VP, Von Gadow K (1997). Fitting a bivariate distribution to diameter-height data of forest trees. Indian Forester 123: 815-820. [online] URL: http://www.indianforester.co.in/index.php/indianforester/article/view/6077
Tewari VP, Von Gadow K (1999). Modelling the relationship between tree diameters and heights using SBB distribution. Forest Ecology and Management 119: 171-176. - doi: 10.1016/S0378-1127(98)00520-9
Wang M, Rennolls K (2005). Tree diameter distribution modeling: introducing the logit-logistic distribution. Canadian Journal of Forest Research 35: 1305-1313. - doi: 10.1139/x05-057
Wang M, Rennolls K (2007). Bivariate distribution modeling with tree diameter and height data. Forest Science 53 (1): 16-24. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2007/00000053/00000001/art00002
Watt P, Meredith A, Yang C, Watt MS (2013). Development of regional models of *Pinus radiata* height from GIS spatial data supported with supplementary satellite imagery. New Zealand Journal of Forestry Science 43: 11. [online] URL: http://www.biomedcentral.com/content/pdf/1179-5395-43-11.pdf
Weibull W (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanics 18 (3): 293-297.
Willemse WJ, Kaas R (2007). Rational reconstruction of frailty-based mortality models by a generalisation of Gompertz' law of mortality. Insurance: Mathematics and Economics 40 (3): 468-484. - doi: 10.1016/j.insmatheco.2006.07.003
Zhang L, Packard KC, 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. Canadian Journal of Forest Research 33: 1340-1347. - doi: 10.1139/x03-054
Zucchini W, Schmidt M, Von Gadow K (2001). A model for the diameter-height distribution in an uneven-aged beech forest and a method to assess the fit of such models. Silva Fennica 35 (2): 169-183. - doi: 10.14214/sf.594

