UDC 630*228.7: 630*521.1/522.31

APPLICATION OF EXTREME VALUE DISTRIBUTION FOR ASSIGNING OPTIMUM FRACTIONS TO DISTRIBUTIONS WITH BOUNDARY PARAMETERS: AN EUCALYPTUS PLANTATIONS CASE STUDY

F. N. Ogana[1], J. S. A. Osho[1], J. J. Gorgoso-Varela[2]

[1] *University of Ibadan*
Ibadan, Oyo State, 200284 Nigeria

[2] *Föra Forest Technologies Co.*
Calle Eduardo Saavadra, 38, Valladolid, 47001 Spain

E-mail: ogana_fry@yahoo.com, jsaosho@yahoo.com, josejavier.gorgoso@usc.es

Received 24.01.2018

The search for an optimum value to constrain boundary parameters in distribution models can be (and is) laborious and time-consuming. The accuracy of a distribution fit depends on the predetermined values of the boundary parameters. In this study, we applied the extreme value distributions derived from the generalized extreme value (GEV) in assigning the optimum constant to a distribution with boundary parameters. GEV subfamily (type 1), Gumbel's distribution, was used to generate constant values which were used as a fraction of the minimum and maximum diameter and height data. The effectiveness of these values was established using five distribution models: logit-logistic (LL), Burr XII, Dagum, Kumaraswamy, and Johnson's S_B distributions. The distributions were fitted with maximum accuracy to the diameter and height data collected on 90 *Eucalyptus camaldulensis* Dehn sample plots. Model assessment was based on negative log-likelihood (-ΛΛ), Kolmogorov-Smirnov (*K-S*), Cramér-von Mises (W^2), Reynold's error index (*EI*), and mean square error (*MSE*). The result showed that the performance of the distributions was improved, especially for the height distribution, compared to other constant values. Gumbel's distribution can be applied whenever (where) a boundary constraint is to be imposed on the location and scale parameters of the distribution models.

Keywords: *generalized extreme value, Gumbel, Kumaraswamy, logit-logistic, Burr XII, Dagum, Johnson* S_B.

How to cite: *Ogana F. N., Osho J. S. A., Gorgoso-Varela J. J.* Application of extreme value distribution for assigning optimum fractions to distributions with boundary parameters: an eucalyptus plantations case study // *Sibirskij Lesnoj Zurnal* (Sib. J. For. Sci.). 2018. N. 4: 39–48 (in English with Russian abstract).

DOI: 10.15372/SJFS20180405

INTRODUCTION

The concept of extreme value distribution modelling originated from the probability theory and statistics. It is also known as Fisher-Tippett distribution (Willemse and Kaas, 2007). It is a family of continuous probability distributions (Gorgoso-Varela et al., 2015). The generalized extreme value (GEV) distribution integrates Gumbel (type I), Fréchet (type II), and Weibull (type III) distributions into a family of models. GEV has been applied to forecast extreme events, such as earthquakes, floods, and other natural disasters. For example, Feng et al. (2007) applied GEV distribution to model annual extreme precipitation in China. Mitková and Halmová (2014) used Gumbel distribution to model the joint distribution of flood peak discharges, volume and duration for the Danube River in Bratislava. Burke et al. (2010) used Gumbel distribution to describe droughts and projected change for the UK.

Gorgoso-Varela and Rojo-Alboreca (2014) suggested that the extreme value distributions could be applied to forestry study to describe the maximum and minimum diameters of trees. For example, the distributions commonly used for size class modelling (e. g. Weibull, beta, Johnson S_B, and Burr) are characterized by location, scale, and shape parameters. Location is usually related to minimum diameter and minimum height (Zhang et al., 2003; Parresol, 2003; Cao, 2004). How accurately these distributions are fitted depends on predetermined values of the location parameter. Researchers have used different algorithms to select values of the location parameter of the distributions commonly used in forestry studies (Zhang et al., 2003; Parresol, 2003; Scolforo et al., 2003; Gorgoso et al., 2012). Wang and Rennolls (2005, 2007) applied a straightforward maximum likelihood estimation method for this parameter; however, convergence was not achieved for some plots. This is a major drawback of the use of the maximum likelihood estimation method.

The concept of extreme distribution modelling was introduced to forestry by Gorgoso-Varela and Rojo-Alboreca (2014), when they used Gumbel and Weibull functions to model extreme diameter distribution values for forest stands. They concluded that the information on the distributions of minimum diameters could be helpful when choosing the most suitable values of the location parameter. Gorgoso-Varela et al. (2015) evaluated the performance of Gumbel, Fréchet, and Weibull distributions. They found Gumbel and Weibull distributions to be useful for modelling the minimum and maximum tree heights. These distributions appeared to be an important tool of distribution modelling in forestry. Therefore, the main purpose of this study was to examine the effectiveness of the use of extreme value distributions for assigning an optimum constant of a distribution with boundary parameters for modelling tree diameter and height data.

METHODOLOGY

Data. The data for this study were collected in an *Eucalyptus camaldulensis* Dehn plantation growing in the Afaka Forest Reserve, Kaduna State, Nigeria. The plantation is located 10.58°–10.60° N and 7.35°–7.37° E, at an elevation of about 600 m above sea level, and occupies an area of 2700 ha (Ogana et al., 2018) (Fig. 1). It is an experimental plantation site, which was established to reduce deterioration and desertification of the Northern Guinea savannah of Nigeria. The data were obtained on 90 sample plots of 625 m^2 each and covered five tree age series of 7, 26, 27, 28, and 29 years. Among the stand variables computed from the inventory data were: stand density, quadratic mean tree diameter, mean tree height, dominant stand height, and basal area per ha. The statistics are summarized in Table 1.

Extreme distribution modeling. Prior to modelling, the minimum and maximum diameters and heights were extracted from each plot to form the distributions of the extreme values. Then we fitted a theoretical extreme distribution, Gumbel distribu-

Fig. 1. Study site of 29 years old eucalyptus *Eucalyptus camaldulensis* Dehn plantation (sample plot 1) in the Afaka Forest Reserve, Kaduna State, Nigeria.

Table 1. Statistics of stand variables

Stand variable	Statistics			
	Mean	Maximum	Minimum	Standard deviation
Tree diameter, cm	10.28	47.43	2.0	6.19
Tree height, m	12.31	39.60	2.10	6.14
Quadratic mean	11.83	23.95	5.89	3.79
Dominant stand height, m	21.01	30.60	9.00	5.38
Stand density, trees/ha^{-1}	753.24	1328	448.0	202.84
Basal area, $m^2 \cdot ha^{-1}$	8.53	27.39	1.73	4.92

Note: Here and in Table 3 and 4 number of sample plots – 90.

Table 2. The FA test for fitting Gumbel functions to extreme tree diameter and height values

Method	Parameter	*K-S*	W^2	*EI*	*MSE*
Moments	D_{min}	0.1386	0.0974	0.2998	0.0373
	D_{max}	0.0964	2.2070	1.4849	0.0003
	H_{min}	0.0654	0.0299	0.1264	0.0320
	H_{max}	0.1024	2.3881	1.5445	0.0006
Mode	D_{min}	0.4095	0.0996	0.3034	0.0372
	D_{max}	0.1069	1.9627	1.4002	0.0003
	H_{min}	0.2651	0.3537	0.5829	0.0153
	H_{max}	0.3084	28.607	5.3483	0.0006

Note: Here and in Table 3 and 4 *K-S* – Kolmogorov-Smirnov distribution; W^2 – Cramér-von Mises distribution; *EI* – Reynolds' error index; *MSE* – mean square error; D_{min}, H_{min}, D_{max}, and H_{max} – minimum and maximum tree diameter and height.

tion, with the method of moments and mode, to eucalyptus extreme diameter and height values. The values of the scale parameter of Gumbel distribution were used as the constants for the boundary parameters of Johnson's S_B, logit-logistic, Kumaraswamy, Burr XII, and Dagum distributions. Gumbel, Weibull, and Fréchet distributions formed the subfamily of GEV distribution. These distributions are sensitive to extreme values. Information about their parameters, especially the scale parameters, could help choose the location and scale parameters of the distributions. This was the rationale for adopting this approach in the study. GEV distribution details presented at earlier publications of Gorgoso-Varela and Rojo-Alboreca (2014) and Gorgoso-Varela et al. (2015). Gumbel cumulative distribution function (*CDF*), as formulated by Gumbel (1954), is expressed as:

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

where: $-\infty < x\ \infty$, μ is the mode value (location parameter) and β is the scale parameter. The method of moments and mode was used to fit Gumbel distribution.

Table 2 shows the fit accuracy (FA) of the minimum and maximum tree diameters and heights of Gumbel distribution.

As is clear from the statistics, the method of moments was more suitable than that of mode concerning fitting the minimum and maximum tree diameters and heights. With the former method, FA values were relatively small as compared with FA resulted from the mode method. Therefore, the method of moment was chosen.

Boundary constraint. The distributions were classified as those with both lower and upper boundary parameters (S_B, LL and Kum distributions) and those with only the lower boundary parameter (Burr XII and Dagum distributions). The following boundary constraints were used and assessed for performances:

1. Location parameter = H_{min}/D_{min} – 0.5 and range parameter = H_{max}/D_{max} + 0.1

2. Location parameter = H_{min}/D_{min} – GV1 and range parameter = H_{max}/D_{max} + 0.1

3. Location parameter = H_{min}/D_{min} – GV1 and range parameter = D_{max} + GV2, for S_B, LL and Kum distributions.

The lower boundaries for Burr XII and Dagum distributions were constrained by:

1. Location parameter = $H_{min}/D_{min} - 0.5$
2. Location parameter = $H_{min}/D_{min} - \text{GV1}$,

where GV1 = 1.34 and GV2 = 6.79 were the values of Gumbel scale parameters obtained from modelling the minimum and maximum tree diameter distribution. The same constraints were applied to tree height distributions, whereas GV1 = 1.15 and GV2 = 5.39 for the height distribution.

Model specification. We evaluated the efficiency of Gumbel fractions on five distribution models with boundary parameters. The models considered were Burr XII, Burr III, (i. e. Dagum), Kumaraswamy, logit-logistic, and Johnson's S_B distributions. These distributions were characterized by four parameters, namely location, scale, and two shape parameters.

Johnson's S_B distribution. The 4-parameter S_B probability density function (Johnson, 1949) is expressed as:

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

where: $\xi < x < \xi + \lambda$, $-\infty < \xi < +\infty$, $-\infty < \gamma < +\infty$, $\lambda > 0$, and $\delta > 0$. S_B function has location parameter, ξ the scale parameter, λ, and the shape parameters, γ and δ, (asymmetry and kurtosis, respectively).

Logit-logistic (LL) distribution. The univariate logit-logistic distribution introduced to forestry by Wang and Rennolls (2005) and used recently by Gorgoso-Varela et al. (2016) was use in this study. The probability density function (*PDF*) and cumulative distribution function (*CDF*) are given as:

$$f(x) = \frac{\lambda}{\sigma} \frac{1}{(x-\xi)(\xi+\lambda-x)} \times$$
$$\times \frac{1}{e^{-(\mu/\sigma)}\left(\frac{x-\xi}{\xi+\lambda-x}\right)^{1/\sigma} + e^{\mu/\sigma}\left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-(1/\sigma)} + 2}, \quad (3)$$

$$F(x) = \frac{1}{1+e^{\mu/\sigma}\left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-(1/\sigma)}}, \quad (4)$$

where: $f(x)$ is probability density function, $F(x)$ is cumulative distribution function, x is diameter and height, μ = mu and σ = sigma are the shape parameters. Other parameters are previously defined in the equation 1.

Burr XII distribution. Burr (1942) constructed a 4-parameter Burr XII distribution. This distribution was introduced to forestry by Wang and Rennolls (2005) and was recently used by Aigbe and Omokhua (2014). *PDF* and *CDF* are expressed as:

$$f(x) = \frac{\alpha k\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}}{\beta\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}, \quad (5)$$

$$F(x) = \left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-k}, \quad (6)$$

where: $f(x)$ is probability density function (*PDF*); $F(x)$ is cumulative distribution function (*CDF*) k and α are two shape parameters ($k > 0$; $\alpha > 0$); β is scale parameter ($\beta > 0$); and γ is the lower boundary parameter, i. e. location parameter ($\gamma \equiv 0$ yields 3-parameter Bur XII distribution). Burr XII distribution has a closed form of *CDF*.

Dagum (Burr III) distribution. Burr (1942) constructed 4-parameter Burr III distribution. This distribution was first introduced to forest modelling by Lindsay et al. (1996). It is also known as Dagum distribution. To distinguish between Burr III and XII, Dagum distribution was used as a substitution of Burr III distribution in this study. *PDF* and *CDF* are expressed as:

$$f(x) = \frac{\alpha k\left(\frac{x-\gamma}{\beta}\right)^{\alpha k-1}}{\beta\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}, \quad (7)$$

$$F(x) = \left(1+\left(\frac{x-\gamma}{\beta}\right)^{-\alpha}\right)^{-k}. \quad (8)$$

All the parameters are predefined. Burr III distribution has a closed form of *CDF*.

Kumaraswamy distribution (Kum). Kumaraswamy (1980) introduced a 4-parameter distribution based on the results of his study on «a generalized probability density function for double-bounded random processes». It is a continuous probability distribution defined in an interval of 0, 1. Kum distribution had previously not been evaluated for forest modelling, as far as we know. The probability density function $f(x)$ and cumulative distribution function $F(x)$ are expressed as:

$$f(x)=\frac{\alpha_1\alpha_2\left[\frac{x-\xi}{\lambda-\xi}\right]^{\alpha_1-1}\left(1-\left[\frac{x-\xi}{\lambda-\xi}\right]^{\alpha_1}\right)^{\alpha_2-1}}{\lambda-\xi}, \quad (9)$$

$$F(x)=1-\left(1-\left[\frac{x-\xi}{\lambda-\xi}\right]^{\alpha_1}\right)^{\alpha_2}, \quad (10)$$

where: x is diameter/height; α_1, α_2 are shape parameters ($\alpha_1 > 0$, $\alpha_2 > 0$); and ξ, λ are boundary parameters ($\xi < \lambda$). Kum distribution had a closed *CDF*.

Fitting method. The method of maximum likelihood was used to fit S_B, LL, Burr XII, Dagum, and Kum distributions to tree diameter and height data. It involved taking partial derivatives of the log-likelihood function, regarding each of the distribution parameters, and setting the expression equal to zero and then solve it by a numerical iterative algorithm to obtain ML estimates. This was achieved using the «optim function» in R (R Core Team, 2016).

The distributions were evaluated based on a negative log-likelihood criterion (–ΛΛ), since the fitting was done using the maximum likelihood estimation (MLE). It is a deviance statistic (Wang and Rennolls, 2007). Kolmogorov-Smirnov (*K-S*), Cramér-von Mises (W^2), Reynold's error index (*EI*), and mean square error (*MSE*) were also used for the model assessment. The smaller the values of the fit indices, the better is the model.

$$K\text{-}S = \mathrm{Sup}x\,[F(x_i) - F_0(x_i)], \quad (11)$$

where Sup$x$ is the supremum value for x:

$$W^2=\sum_{i=1}^{n}\left\{\hat{F}(x_i)-\frac{(i-0.5)}{n}\right\}^2+\frac{1}{12n}, \quad (12)$$

$$EI=\sum_{i=1}^{n}\left[\hat{F}(x_i)-\frac{(i-0.5)}{n}\right], \quad (13)$$

$$MSE=\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{n}. \quad (14)$$

$F(x_i)$ is the cumulative frequency distribution observed for x_i sample ($i = 1, 2, \ldots, n$), $F_0(x_i)$ is the probability of a theoretical cumulative frequency distribution; Y_i is the observed value, and $\hat{Y}_i$ is the theoretical value predicted by the model.

RESULTS AND DISCUSSION

The assessment of the overall fitting performance of five univariate distributions of interest with respect to the boundary constraints are presented in Table 3 and 4.

The results for the fitted S_B diameter distributions showed that constraint 2 had the smallest mean –ΛΛ value (–131.2065), whereas constraint 1 had the largest value (–129.3). However, constraint 1 had the smallest mean *K-S* value (0.1133). Constraint 3 had the smallest *MSE* value (0.0014). These three constraints produced approximately the same mean values of W^2 and *EI* (Table 3).

Similarly, the smallest mean –ΛΛ (–131.6463) was recorded in constraint 2 for LL. This value increased in constraint 3. Constraint 1 had the largest –ΛΛ. When *K-S* was used as the criterion, constraint 1 had the smallest mean values (0.0998); these values increased in constraint 3. Constraint 2 had the largest mean *K-S* value.

The mean values of W^2 and *EI* for the constraints were approximately the same. The fit of Kum distribution was improved by the use of Gumbel values, with constraint 2 having the smallest mean –ΛΛ value (–136.5602). Constraint 3 had the smallest mean *K-S*, W^2, *EI*, and *MSE* values (0.1276, 0.0831, 0.955, and 0.0016, respectively). As for Burr XII distribution, the two constraints had the same –ΛΛ value (–129.7471). However, constraint 1 had smaller mean *K-S*, W^2, *EI*, and *MSE* values (0.1059, 0.00745, 0.8871, and 0.0014, respectively). The result for Dagum distribution showed that constraint 1 had smaller values for all the fit indices, except log-likelihood (–ΛΛ). Its mean –ΛΛ, *K-S*, W^2, *EI*, and *MSE* values were –129.8797, 0.1013, 0.0638, 0.8117, and 0.0013, respectively. Constraint 2 had a smaller –ΛΛ value (–129.877).

The results for the fitted S_B height distribution showed that Gumbel values improved the fitting performance of the distribution. Constraint 3 had the smallest values for most of the fit indices. Its –ΛΛ, *K-S*, W^2, *EI*, and *MSE* values were –135.601, 0.098, 0.0513, 0.7456, and 0.0013, respectively (Table 4).

These values increased proceeding to constraint 2 and then to constraint 1. Constraint 3 had the smallest mean *K-S*, W^2, *EI*, and *MSE* values (0.0843, 0.0402, 0.6718, and 0.0013, respectively) for LL marginal distribution. Only in –ΛΛ was constraint 2 slightly ranked higher than constraint 3. Constraint 3 gave the smallest mean *K-S*, W^2, *EI*, and *MSE* (0.1019, 0.0512, 0.7549, and 0.0013, respectively) for Kum tree height distribution. However, the log-likelihood (–ΛΛ) values of con-

Table 3. Mean goodness-of-fit test for the fitted diameter distributions

Distribution	Parameter		–ΛΛ	*K-S*	W^2	*EI*	*MSE*
	ξ	λ					
S_B	D_{min} – 0.5	D_{max} + 0.1	–129.3000	0.1133	0.0772	0.9059	0.0015
	D_{min} – GV1	D_{max} + 0.1	–131.2065	0.1199	0.0765	0.9119	0.0015
	D_{min} – GV1	D_{max} + GV2	–130.4812	0.1157	0.0775	0.9082	0.0014
LL	D_{min} – 0.5	D_{max} + 0.1	–129.1161	0.0998	0.0617	0.8405	0.0015
	D_{min} – GV1	D_{max} + 0.1	–131.6463	0.1033	0.0625	0.8491	0.0015
	D_{min} – GV1	D_{max} + GV2	–130.8504	0.1012	0.0657	0.8669	0.0014
Kum	D_{min} – 0.5	D_{max} + 0.1	–133.3320	0.1484	0.1248	1.2137	0.0018
	D_{min} – GV1	D_{max} + 0.1	–136.5602	0.1598	0.1453	1.3237	0.0018
	D_{min} – GV1	D_{max} + GV2	–132.8000	0.1276	0.0831	0.9550	0.0016
Burr XII	D_{min} – 0.5		–129.7471	0.1059	0.0745	0.8871	0.0014
	D_{min} – GV1		–129.7471	0.1682	0.1419	1.2840	0.0018
Dagum	D_{min} – 0.5		–128.8797	0.1013	0.0638	0.8117	0.0013
	D_{min} – GV1		–129.8770	0.1548	0.1203	1.1909	0.0017

Note: –ΛΛ – negative log-likelihood; GV1 – Gumbel value 1 = 1.34; GV2 – Gumbel value 2 = 6.79.

Table 4. Mean goodness-of-fit test for the fitted height distributions

Distribution	Parameter		–ΛΛ	*K-S*	W^2	*EI*	*MSE*
	ξ	λ					
S_B	H_{min}–0.5	H_{max} + 0.1	–135.3756	0.1025	0.0558	0.7836	0.0014
	H_{min} – GV1	H_{max} + 0.1	–135.5972	0.0987	0.0499	0.7432	0.0013
	H_{min} – GV1	H_{max}+GV2	–135.6010	0.0980	0.0513	0.7456	0.0013
LL	H_{min} – 0.5	H_{max} + 0.1	–134.6366	0.0893	0.0437	0.7059	0.0014
	H_{min} – GV1	H_{max} + 0.1	–135.9805	0.0894	0.0435	0.7053	0.0014
	H_{min} – GV1	H_{max} + GV2	–135.3591	0.0843	0.0402	0.6718	0.0013
Kum	H_{min} – 0.5	H_{max} + 0.1	–137.4682	0.1281	0.0896	1.0486	0.0015
	H_{min} – GV1	H_{max} + 0.1	–139.1868	0.1350	0.1002	1.1129	0.0015
	H_{min} – GV1	H_{max} + GV2	–136.3794	0.1019	0.0512	0.7549	0.0013
Burr XII	H_{min} – 0.5		–135.3452	0.0927	0.0495	0.7153	0.0013
	H_{min} – GV1		–135.5400	0.0886	0.0451	0.6851	0.0013
Dagum	H_{min} – 0.5		–134.6450	0.0876	0.0401	0.6351	0.0012
	H_{min} – GV1		–135.5258	0.0853	0.0426	0.6569	0.0012

Note: GV1 – Gumbel value 1 = 1.15; GV2 – Gumbel value 2 = 5.39.

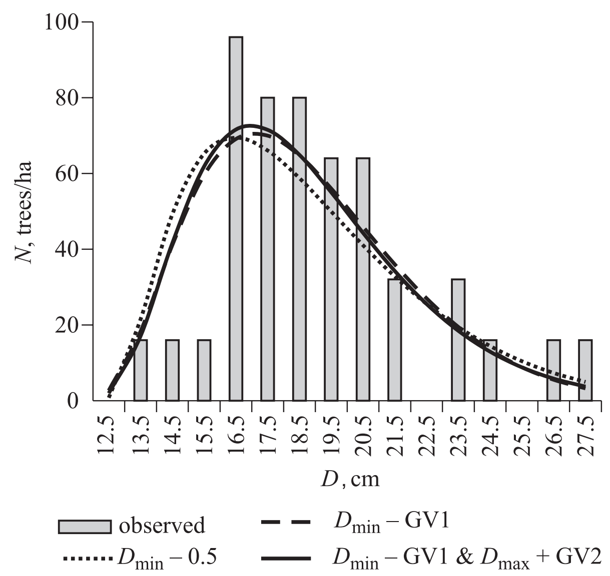

Fig. 2. Observed diameter distributions, Johnson S_B fitted using the three boundary constraints for plot 1 with a stand age of 29 years.

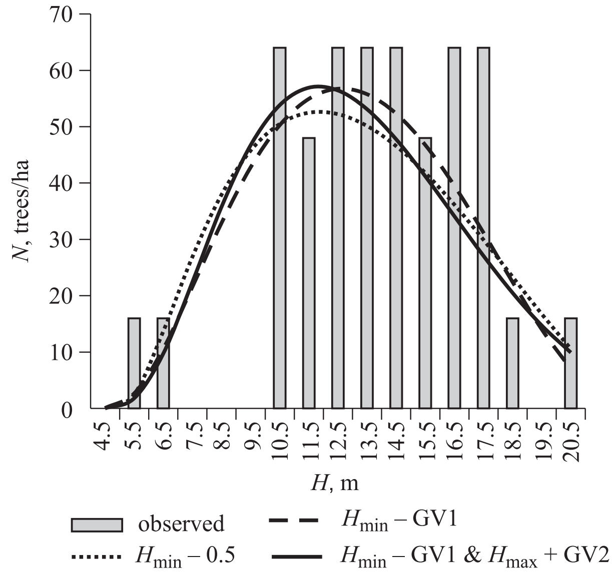

Fig. 3. Observed height distributions, Johnson S_B fitted using the three boundary constraints for plot 1 with a stand age of 29 years.

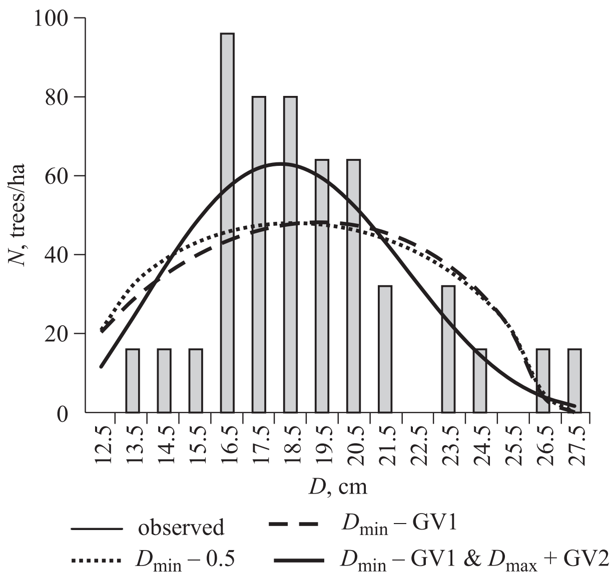

Fig. 4. Observed diameter distributions, Kum fitted using the three boundary constraints for plot 1 with a stand age of 29 years.

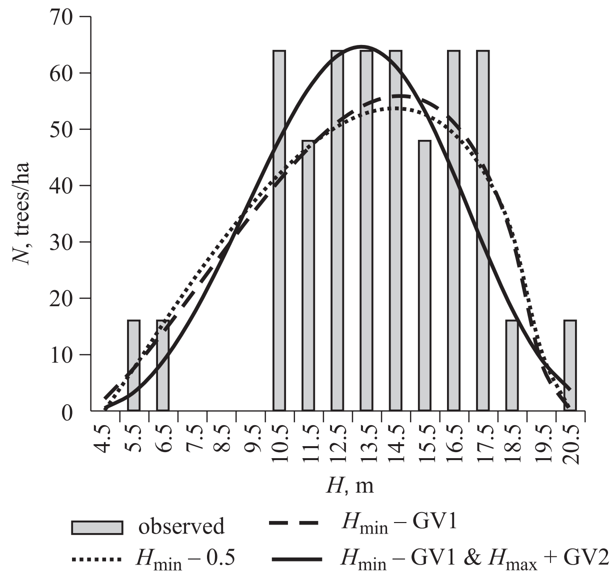

Fig. 5. Observed height distributions, Kum fitted using the three boundary constraints for plot 1 with a stand age of 29 years.

straint 2 and constraint 1 were smaller than those of constraint 3. For Burr XII distribution with just two constraints, the results showed that constraint 2 (H_{min} – GV1) had smaller values for the fit statistics. Its –ΛΛ, *K-S*, W^2, *EI*, and *MSE* were –135.54, 0.0886, 0.0451, 0.6851, and 0.0013, respectively. The results for Dagum distribution showed that constraint 2 had smaller values for –ΛΛ and *K-S*, (135.5258 and 0.0853, respectively), which were the two most important fit statistics. Constraint 1 had smaller mean W^2 and *EI* values (0.0401 and 0.6351, respectively). Both constraints had the same *MSE* value of 0.0012. Generally, the results of Gumbel fractions were comparable to the optimum fraction (D_{min} /H_{min} – 0.5), and in some cases the fractions performed even better.

The graphs of the diameter and height distributions resulting from the use of the three boundary constraints are presented in Fig. 2–5.

The graphs showed the number of trees in each diameter and height class per ha. The shape of the observed diameter and height distributions were mainly right-skewed asymmetrical, i. e. stretched-tailed to the right rather than symmetrical, except in few plots.

Only the graph of the most common and much used Johnson's S_B and newly introduced Kum distributions are presented here using a representative sample plot (plot 1 with the trees 29 years old). The trees on this plot had an average diameter and height of 18.9 cm and 13.2 m, respectively, with the stand density being 544 trees/ha. As is clear from the graph, Gumbel scale values improved the performance of the distribution compared to the usual constant (i. e. $D_{min} – 0.5$). This was also obvious for the fitted Johnson S_B height distribution (Fig. 3) and the fitted Kum diameter and height distributions (Fig. 4 and 5).

In this study, a new approach was applied for selecting the optimum values for the lower boundary, i. e. location parameters, and of upper boundary parameters of Johnson's S_B, logit-logistic (LL), Kum, Bur XII, and Dagum distributions. These fractions were generated from Gumbel scale parameter fitted to the extremes values of tree diameter and height of eucalyptus stands. The rationale for this approach was to find optimum values for the location and scale parameters that is stand specific, rather than using arbitrarily selected fractions. Only the scale parameter of Gumbel function was used. The values were 1.34 and 1.15 for minimum diameter and minimum height, respectively; 6.79 and 5.39 for maximum diameter and maximum height, respectively. These fractions were compared with the best fractions reported in quantitative forestry (e. g. min, 0.1, 0.5 etc.). We also tried to use fractions of Weibull scale parameter, but their performance was poor; as such, it was not documented in this study.

All the fractions used in this study performed generally well. Moreover, the performances of S_B, LL, Kum, Burr XII and Dagum distributions were slightly improved, especially for the height distributions, when Gumbel scale value was used. No convergence was achieved when the location and scale parameters were set to minimum and maximum diameter and height. The absence of convergence allowed us to conclude that distribution accuracy depended on the predetermined values of the location parameter and, to some extent, of the scale parameter. This effect varied with distribution, estimation method, and tree species. For example, the performance of the distributions improved as the magnitude of the fractions increased (min < 0.1 < 0.5 < Gumbel scale) from using the fit indices applied in this study. However, when the magnitude of a fraction was further increased by using Weibull scale with a large value as against Gumbel, the performance of the diameter and height distributions was poor (though it was not documented). A similar trend was observed by Zhang et al. (2003), who compared different estimation methods to fit Weibull and S_B distributions to mixed spruce-fir stands. The location parameter was assumed to equal ($D_{min} – 0.5$) and different values (0.5, 1.0, 1.5 and 2.0) were tried. Their results showed that Reynolds' error index increased with increasing constants for S_B and Weibull distributions. The constraint ($D_{min} – 0.5$) used in our study also performed well based on negative log-likelihood, Kolmogorov-Smirnov, Cramér-von Mises, mean square error, and Reynolds' error index. However, D_{min} –Gumbel scale performed better, especially for tree height distributions.

Furthermore, this Gumbel fraction provided an alternative way of fitting height distributions with S_B Knoebel and Burkhart (KB) method. For example, normal KB fractions (location = $D_{min} – 1.3$; scale = D_{max} – location + 3.8) were derived for the diameter distribution by Knoebel and Burkhart (1991). These constants (1.3 and 3.8) are known to be functions of plot size, stand structure, and even tree species. With the direct use of the fraction for characterizing height distribution, the results obtained were not better than with Gumbel fractions. Zhou and McTague (1996) applied a complex algorithm to extrapolate height fractions equivalent to diameter fractions, so that consistency with KB method could be maintained. The authors fitted a simple linear regression to tree height-diameter pairs for each plot. They predicted height for at $D = D_{min}$ and at $D = D_{min} – 1.3$, subtracted the difference in height from these two predictions, and replaced D_{min} and 1.3 by H_{min} and the difference to obtain the location parameter for the height distribution. They also predicted height at $D = D_{max}$ and at $D = D_{max} + 3.8$; subtracted the difference in height from these predictions, and replaced D_{max} and 3.8 by H_{max} and the difference to obtain the scale parameter for the height distribution. Although this seems to be a good approach, it is tedious and unnecessary.

Li et al. (2002) used KB method with predetermined values for the location and scale parameters of S_B to fit the marginal diameter and height distributions of a Douglas fir stand. The minimum tree height was used for the location parameter and a range of heights for the scale parameter. They reported poor performance with S_B as compared to the generalized beta distribution. However, Wang (2005) used predetermined constants (1.3 and 5.1) still in the sense of KB method to obtain a better result for S_B and LL diameter distributions. The rationale for adopting these constants was not clear from the paper. The author observed that these constraints improved the cumulative distribution function-based least square method (*CDF*-based LS) more than the maximum likelihood method. Nevertheless, both methods produced satisfactory results in his study. No study has been done on the use of a boundary constraint for Burr XII, Dagum and Kum distributions.

Application of boundary constrains to the location and sometimes to scale parameters of distribution functions is a common practice in quantitative forestry, because the accuracy of the results of using distribution functions in size class modelling depends on predetermined values of these parameters. In the absence of a boundary constrain for the location parameter, the estimate could sometimes be biologically unreasonable, especially when MLE is used for fitting the models, for example, where an estimate for the location parameter is given as a negative value. This is not possible concerning diameter and/or height distributions. The location parameter marks the beginning of the diameter distribution, which is the minimum tree diameter observed. The location parameter could be zero, especially when tree height is 1.3 meters. Trees 1.3 m high are rarely measured, unless it is a young forest stand.

CONCLUSION

To sum up, we assessed the effectiveness of Gumbel scale as an optimum constant for improving the fitting performances of S_B, LL, Burr XII, Dagum and Kum distributions. The results were quite satisfactory. The use of Gumbel scale value improved the performance of some of the distributions, especially of the height distribution. The implication is that, it will reduce the time required for testing different constant values during the model fitting process.

REFERENCES

Aigbe H. I., Omokhua G. E. Modeling diameter distribution of the tropical rainforest in Oban Forest Reserve // J. Environ. Ecol. 2014. V. 5. N. 2. P. 130–143.

Burke E. J., Perry R. H. J., Brown S. J. An extreme value analysis of UK drought and projections of change in the future // J. Hydrol. 2010. V. 388. Iss. 1–2. P. 131–143.

Burr I. W. Cumulative frequency functions // Ann. Math. Statist. 1942. V. 13. N. 2. P. 215–232.

Cao Q. V. Predicting parameters of a Weibull function for modelling diameter distribution // For. Sci. 2004. V. 50. Iss. 5. P. 682–685.

Feng S., Nadarajah S., Hu Q. Modelling annual extreme precipitation in China using the generalized extreme value distribution // J. Meteorol. Soc. Jap. 2007. V. 85. N. 5. P. 599–613.

Gorgoso-Varela J. J., García-Villabrille J. D., Rojo-Alboreca A. Modeling extreme values for height distributions in *Pinus pinaster*, *Pinus radiata* and *Eucalyptus globulus* stands in northwestern Spain // iForest. 2015. V. 9. P. 23–29.

Gorgoso-Varela J. J., García-Villabrille J. D., Rojo-Alboreca A., Gadow K. V., Álvarez-González J. G. Comparing Joshnson's S_{BB}, Weibull and Logit-Logistic bivariate distributions for modelling tree diameters and heights using copulas // For. Syst. 2016. V. 25. N. 1. P. 1–5.

Gorgoso-Varela J. J., Rojo-Alboreca A. Use of Gumbel and Weibull functions to model extreme values of diameter distributions in forest stands // Ann. For. Sci. 2014. V. 71. Iss. 7. P. 741–750.

Gorgoso J. J., Rojo-Alboreca A., Camara-Obregon A., Dieguez-Aranda U. 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 // For. Syst. 2012. V. 21. N. 3. P. 446–459.

Gumbel E. J. Statistical theory of extreme values and some practical applications: A series of lectures. U. S. Nat. Bureau of Standards. Appl. Math. Ser., 33. Washington, D.C., U. S. Govt. Print. Office, 1954. 51 p.

Johnson N. L. Systems of frequency curves generated by methods of translation // Biometrika. 1949. V. 36. N. 1–2. P. 149–176.

Knoebel B. R., Burkhart H. E. A bivariate distribution approach to modeling forest diameter distributions at two points in time // Biometrics. 1991. V. 47. N. 1. P. 241–253.

Kumaraswamy P. A generalized probability density function for double-bounded random processes // J. Hydrol. 1980. V. 46. Iss. 1–2. P. 79–88.

Li F., Zhang L., Davis C. J. Modelling the joint distribution of tree diameters and heights by bivariate generalized beta distribution // For. Sci. 2002. V. 48. Iss. 1. P. 47–58.

Lindsay S. R., Wood G. R., Woollons R. C. Modelling the diameter distribution of forest stands using the Burr distribution // J. Appl. Statist. 1996. V. 23. Iss. 6. P. 609–620.

Mitková V. B., Halmová D. Joint modelling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava // J. Hydrol. Hydromech. 2014. V. 62. N. 3. P. 186–196.

Ogana F. N., Osho J. S. A., Gorgoso-Varela J. J. An approach to modeling the joint distribution of tree diameter and height data // J. Sustain. For. 2018. V. 37. Iss.5 P. 475–488.

Parresol B. R. Recovering parameters of Johnson's S_B distribution. Res. Pap. SRS-31. Asheville, NC: USDA, For. Serv., South. Res. St., 2003. 9 p.

R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. http://www.R-project.org/

Scolforo J. R. S., Tabai F. C. V., Macedo R. L. G. d., Acerbi F. W., Assis A. L. d. S_B distribution's accuracy to represent the diameter distribution of *Pinus taeda*, through five fitting methods // For. Ecol. Manag. 2003. V. 175. Iss. 1–3. P. 489–496.

Wang M. Distributional modelling in forestry and remote sensing. PhD thesis, Univ. Greenwich, Greenwich, UK, 2005. 187 p.

Wang M., Rennolls K. Tree diameter distribution modelling: introducing the logit-logistic distribution // Can. J. For. Res. 2005. V. 35. N. 6. P. 1305–1313.

Wang M., Rennolls K. Bivariate distribution modeling with tree diameter and height data // For. Sci. 2007. V. 53. Iss. 1. P. 16–24.

Willemse W. I., Kaas R. Rational reconstruction of frailty-based mortality models by generalisation of Gompertz' law of mortality // Insurance: Math. Econ. 2007. V. 40. Iss. 3. P. 468–484.

Zhang L., Packard K. C., Liu C. 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. 2003. V. 33. N. 7. P. 1340–1347.

Zhou B., McTague J. P. Comparison and evaluation of 5 methods of estimation of the Johnson system parameters // Can. J. For. Res. 1996. V. 26. N. 6. P. 928–935.

УДК 630*228.7: 630*521.1/522.31

ПРИМЕНЕНИЕ РАСПРЕДЕЛЕНИЯ ЭКСТРЕМАЛЬНЫХ ЗНАЧЕНИЙ ДЛЯ ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНЫХ ВЕЛИЧИН РАСПРЕДЕЛЕНИЙ С ГРАНИЧНЫМИ ПАРАМЕТРАМИ НА ПРИМЕРЕ ПЛАНТАЦИЙ ЭВКАЛИПТА

Ф. Н. Огана[1], Дж. С. А. Ошо[1], Х. Х. Горгозо-Варела[2]

[1] *Ибаданский университет*
Нигерия, 200284, Ойо, Ибадан

[2] *Föra Forest Technologies Co.*
Испания, 42004, Сория, ул. Калле Эдуардо Сааведра, 38

E-mail: ogana_fry@yahoo.com, jsaosho@yahoo.com, josejavier.gorgoso@usc.es

Поступила в редакцию 24.01.2018 г.

Поиск оптимальных значений для определения граничных параметров в моделях распределения часто бывает время- и трудоемким. Точность распределения зависит от заданных значений граничных параметров. В этом исследовании мы применили распределения экстремальных значений, полученные из обобщенного экстремального значения (ОЭЗ) при определении оптимальной константы для распределения с граничными параметрами. Для определения постоянных значений использовалось подсемейство ОЭЗ (тип 1) распределение Гумбеля, которое включало данные минимальных и максимальных диаметров и высот деревьев. Эффективность значений оценивалась пятью моделями распределений: логистической, Бурра XII, Дагума, Кумарасвами и Джонсона $С_В$. Функции распределения были подобраны по принципу максимального подобия по данным обмеров диаметров и высот деревьев, полученных на 90 пробных площадях, заложенных в насаждениях (плантациях) эвкалипта камальдульского *Eucalyptus camaldulensis* Dehn. Оценка моделей проводилась с использованием отрицательной логарифмической вероятности (-ΛΛ), критериев Колмогорова-Смирнова (*K-S*), Крамера фон Мизеса (W^2), индекса ошибок Рейнольдса (*EI*) и средней квадратичной ошибки (*MSE*). Результат исследований показал, что производительность распределений была улучшена, особенно для распределения по высоте, по сравнению с другими постоянными значениями. Распределение Гумбеля может применяться всякий раз, когда устанавливается ограничение на параметры местоположения и масштаба моделей распределения.

Ключевые слова: *обобщенное экстремальное значение, Гумбель, Кумарасвами, логистическое распределение, Бурр XII, Дагум, Джонсон $С_В$.*