Regular Article
pISSN: 2288-9744, eISSN: 2288-9752
Journal of Forest and Environmental Science
Vol. 36, No. 1, pp. 7-16, March, 2020
https://doi.org/10.7747/JFES.2020.36.1.7

JFES
Journal of Forest and Environmental Science

Tree Size Distribution Modelling: Moving from Complexity to Finite Mixture

Friday Nwabueze Ogana[1,*], Onyekachi Chukwu[1] and Samuel Ajayi[2]
[1]*Department of Social and Environmental Forestry, University of Ibadan, Ibadan 200284, Nigeria*
[2]*Department of Forestry and Wildlife Resources Management, Cross River University of Technology, Obubra 5511114, Nigeria*

Abstract

Tree size distribution modelling is an integral part of forest management. Most distribution yield systems rely on some flexible probability models. In this study, a simple finite mixture of two components two-parameter Weibull distribution was compared with complex four-parameter distributions in terms of their fitness to predict tree size distribution of teak (*Tectona grandis* Linn f) plantations. Also, a system of equation was developed using Seemingly Unrelated Regression wherein the size distributions of the stand were predicted. Generalized beta, Johnson's SB, Logit-Logistic and generalized Weibull distributions were the four-parameter distributions considered. The Kolmogorov-Smirnov test and negative log-likelihood value were used to assess the distributions. The results show that the simple finite mixture outperformed the four-parameter distributions especially in stands that are bimodal and heavily skewed. Twelve models were developed in the system of equation-one for predicting mean diameter, seven for predicting percentiles and four for predicting the parameters of the finite mixture distribution. Predictions from the system of equation are reasonable and compare well with observed distributions of the stand. This simplified mixture would allow for wider application in distribution modelling and can also be integrated as component model in stand density management diagram.

Key Words: finite mixture, four-parameter distributions, teak, seemingly unrelated regression, Weibull distribution

Introduction

Modelling tree size distribution is an important aspect of forest growth and yield studies. It is a determinant of stand structure, stability, volume and volume of different products (van Laar and Akça 2007; Gorgoso-Varela and Rojo-Alboreca 2014). Tree size distribution modelling has been an integral part of forest management, planning and research. It has often been used to bridge the gap between stand level models and tree level models (Larsen and Cao 2006). Since the early work of De Liocourt (1898), several probability density functions (pdf) have been used for growth and yield studies with different level of success including gamma, normal, lognormal, Weibull, generalized beta, Johnson's S_B, Logit-Logistic, generalized Weibull distributions (Zhang et al. 2003; Wang and Rennolls 2005; Palahí et al. 2007; Ajayi 2013; Poudel and Cao 2013; Gorgoso-Varela and Rojo-Alboreca 2014; Ogana et al. 2018).

The choice of distribution model in growth and yield studies depends on its relative flexibility (approximate wide varieties of shapes), ease of parameter estimation and simplicity in computing proportion of trees in different classes (Burkart and Tomé 2012). Tree growth characteristics such as diameter and height exhibit different structures due to silvicultural treatments (below or above thinning) and/or

Received: April 11, 2019. Revised: December 11, 2019. Accepted: December 12, 2019.

Corresponding author: Friday Nwabueze Ogana

Department of Social and Environmental Forestry, University of Ibadan, Ibadan 200284, Nigeria
Tel: +234-9055337571, Fax:+234-7037679328, E-mail: ogana_fry@yahoo.com

mortality. According to van Laar and Akça (2007), the structure of a stand during the early stage of development before thinning is symmetrical and can be considered as a normal distribution; but the stand may become skewed (i.e. asymmetric) with increasing age due to mortality or thinning. Thus, because of the effect of thinning and mortality, foresters have depended on the use of complex and more flexible distributions such as the four parameters distributions (e.g. generalized beta, Johnson's SB, Logit-Logistic, generalized Weibull distributions, etc.) for tree size distribution modelling. These distributions are flexible and therefore can describe wide range of shapes (Wang 2005; Wang and Rennolls 2005). However, they require great deal of computations and sometimes may not approximate the stand structure, especially a stand that is bimodal or heavily skewed.

Finite mixture (FM) was introduced to forestry as an alternative technique for describing bimodal or heavily skewed size distributions (Zhang et al. 2001; Liu et al. 2002; Zasada and Cieszewski 2005). Since then several researchers have used FM to classify and/or describe the structure of forest stands (e.g., Zhang and Liu 2006; Jaworski and Podlaski 2012; Liu et al. 2014; Tsogt and Lin 2012; Lin et al. 2016; Podlaski 2017). Recently, Ogana (2018) applied FM to describe the structure of degraded *Gmelina arborea* Roxb stands in Nigeria. The number of components in the mixture distribution and the mixing proportion could either be predetermined (Zasada and Cieszewski 2005) or iteratively searched which can be achieved by maximum likelihood through expectation maximization algorithm.

Generally, yield systems predict values for the parameters of distribution functions (Clutter et al. 1983). The parameters are related either directly with stand variables (parameter prediction model) or diameter moment/percentiles derived from forest stand variables (parameter recovery model). This can be used for implicit prediction of future yield. Despite the flexibility of FM, the application of the distribution to growth and yield models has been limited. To date, no published forestry literature exists in Nigeria that relate the parameters of FM distribution with stand variables as much as we know. Therefore, the objectives of this study were to: compare the complex commonly used four-parameter distributions with a simple finite mixture distribution, and to model the parameters of the finite mixture distribution with stand variables for yield predictions.

Materials and Methods

Data

The study site selected for the study was the *Tectona grandis* Lnn.f (teak) plantation in Omo Forest Reserve, Ogun State, Nigeria. It lies between Latitudes 6°35′ to

Table 1. Descriptive statistics of tree and stand variables of *T. grandis*

Variable	Descriptive statistic			
	Mean	Max	Min	Standard deviation
TREE				
DBH (cm)	17.9	37.9	6.0	5.35
Height (m)	16.7	26.5	6.6	3.51
STAND				
Age (years)	19.6	27.0	13.0	5.28
Quadratic mean (cm)	18.4	22.4	14.3	2.23
DGM (cm)	20.0	23.9	15.3	2.42
Hd (m)	21.3	24.0	17.9	1.87
G (m^2/ha)	24.02	50.78	11.54	8.84
N (trees/ha)	880	1,744	624	196.08
GS (m)	3.4	4.0	2.4	0.30

DGM, basal of central diameter; Hd, dominant height; G, basal area; N, number of trees; GS, growing space.

7°03′N and Longitudes 4°09′ to 4°40′E and occupies an area of 139,100 ha. The data consist of 1919 trees measured from 35 sample plots of 625 m^2 size. Diameter at breast height (DBH at 1.3m above the ground level) and total height were measured with diameter tape and hypsometer to the nearest 0.1 cm, and 0.1 m, respectively. Stand variables including number of trees per ha, quadratic mean diameter, basal area of central diameter, dominant height, growing space and basal area per ha were computed. The descriptive statistics of the data set are presented in Table 1.

Four-parameter distribution models

Four commonly used 4-parameter distributions in forestry were considered in this study including Johnson's S_B (SB), generalized beta (GB), Logit-Logistic (LL) and Generalized Weibull (GW). These distributions are relatively flexible to describe both positive and negative skew distributions (Wang and Rennolls 2005). The four 4-parameter distributions were compared with a simple finite mixture distribution. Tree diameter and height were the tree size variables used because they are the fundamental variables from which other stand variables are derived (Sharma et al. 2018).

The Johnson's S_B distribution: the four parameter Johnson's S_B (Johnson 1949) is given by:

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

Where ξ is the location parameter, λ is the scale parameter, and γ and δ are the shape parameters (i.e., asymmetry and kurtosis parameters, respectively); $\xi < x < \xi + \lambda$, $-\infty < \xi < +\infty$, $-\infty < \gamma < +\infty$, $\lambda > 0$, and $\delta > 0$. Johnson's S_B has no closed-form cumulative distribution function (cdf) and as such, numerical integration was applied. The S_B distribution was recently used by Gorgoso-Varela and Rojo-Alboreca (2014), Ogana et al. (2017) and Ogana et al. (2018).

The Logit-Logistic (LL) distribution: the probability density function (pdf) and cumulative distribution function (cdf) of LL are expressed as:

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

$$F(x) = \frac{1}{1+e^{\mu/\sigma}\left(\frac{x-\xi}{\xi+\lambda-x}\right)^{-(1/\sigma)}} \quad \text{Eq. (3)}$$

Where $f(x)$=probability density function, $F(x)$=cumulative distribution function, x=diameter/height. The parameters μ = mu and σ=sigma are the shape parameters. Other parameters are previously defined in equation (1). This distribution was used by Wang and Rennolls (2005), Gorgoso-Varela et al. (2016) and Ogana et al. (2018)

The generalized Weibull (GW) distribution: this is a four-parameter Weibull distribution introduced to forestry by Wang and Rennolls (2005). Its pdf and cdf are given by:

$$f(x) = \frac{ck}{b}\left(\frac{x-a}{b}\right)^{c-1}\left(e\left[-\left(\frac{x-a}{b}\right)^c\right]\right)\left(1-e\left[-\left(\frac{x-a}{b}\right)^c\right]\right)^{k-1} \quad \text{Eq. (4)}$$

$$F(x) = \left(1-e\left[-\left(\frac{x-a}{b}\right)^c\right]\right)^k \quad \text{Eq. (5)}$$

Where c=shape parameter ($c>0$); k=exponentiated shape parameter ($k>0$); b=scale parameter ($b>0$); a= location parameter. The GW distribution was used by Wang and Rennolls (2005).

The generalized beta (GB) distribution: the pdf of GBD is expressed as:

$$f(x) = \frac{1}{(b-a)\,B(\alpha_1,\alpha_2)}\left(\frac{x-a}{b-a}\right)^{\alpha_1-1}\left(1-\frac{x-a}{b-a}\right)^{\alpha_2-1} \quad \text{Eq. (6)}$$

$$B(\alpha_1,\alpha_2) = \frac{r(\alpha_1)r(\alpha_2)}{r(\alpha_1+\alpha_2)} \quad \text{Eq. (7)}$$

Where α_1, α_2 are the two shape parameters; a and b are the location and scale parameters, respectively, Γ is the gamma function. $a \le x \le b$, $\alpha_1 > 0$, $\alpha_2 > 0$. Just as the S_B distribution, GB has no closed-form. Numerical integration was used to compute class frequency. The GB distribution was used by Li et al. (2002), Wang and Rennolls

(2005) and Jin et al. (2013).

Finite mixture (FM) distribution

The finite mixture distribution for a random continuous variable x (e.g. tree diameter, height etc.) can be expressed as:

$$f(x,\pi)=\sum_{j=1}^{k}\pi_i f_i(x)=\pi_1 f_1(x)+\pi_2 f_2(x)+\cdots+\pi_k f_k(x)$$

Eq. (8)

Where $f_j(x)$=probability density function (pdf) of the jth individual component distribution, π=mixing proportion of the individual components in the mixture distribution which must satisfy the condition: $0\leq\pi_j\leq 1$ and $\pi_k=\sum_{j=1}^{k}\pi_j$. The number of components in the mixture could be predetermined (Zasada and Cieszewski 2005) or search iteratively (Ogana 2018). However, the chance of achieving convergence becomes slimmer when the components in the mixture increases (Jaworski and Podlaski 2012). In this study, a simple scenario of mixing proportion was formulated. Two 2-parameter Weibull function was used as the components of the mixture distribution with a mixing proportion of 0.5. The Weibull function (Weibull, 1951) is the most widely used distribution in tree size modelling because of its relative flexibility, simplicity of model expression and ease of computation. The pdf and cdf of the Weibull mixture distribution are expressed as:

$$f(x)=\pi_1\left(\frac{\alpha_1}{\beta_1}\left(\frac{x}{\beta_1}\right)^{\alpha_1-1}\exp\left[-\left(\frac{x}{\beta_1}\right)^{\alpha_1}\right]\right)+\pi_2\left(\frac{\alpha_2}{\beta_2}\left(\frac{x}{\beta_2}\right)^{\alpha_1-1}\exp\left[-\left(\frac{x}{\beta_2}\right)^{\alpha_2}\right]\right)$$

Eq. (9)

$$F(x)=\pi_1\left(1-\exp\left[-\left(\frac{x}{\beta_1}\right)^{\alpha_1}\right]\right)+\pi_2\left(1-\exp\left[-\left(\frac{x}{\beta_2}\right)^{\alpha_2}\right]\right)$$

Eq. (10)

Where α_1, β_1=shape and scale parameters of the first component in the mixture distribution, α_2 and β_2=shape and scale parameters of the second component. Since the mixing proportion (π) was constrained to be 0.5, thus, the Weibull mixture distribution has four parameters.

Fitting and evaluation

The method of maximum likelihood (MLE) was used to fit the 4-parameter distributions and the finite mixture distributions. MLE involves taking the partial derivatives of the log-likelihood function with respect to the parameters of each distribution, equating the expression to zero and then solved by iterative algorithm. The location and range parameters were constrained to be minimum diameter/height minus 0.5 and maximum diameter/height plus 0.1. Similar constrain was used by Gorgoso-Varela et al. (2016). The distributions were fitted to the diameter and height data from each plot. The distributions were evaluated based on Kolmogorov-Smirnov (Dn) and negative loglikelihood values. The smaller the values of the fit indices, the better the distribution.

$$D_n=Supx\left|F(x_i)-F_0(x_i)\right| \quad \text{Eq. (11)}$$

Where sup$x$ is the supremum value for x; $F(x_i)$ is the cumulative frequency distribution observed for the sample x_i (i=1, 2, ..., n), $F_0(x_i)$ is the probability of the theoretical cumulative frequency distribution. Analysis was carried out in R (R Core Team 2017).

Modelling the diameter distribution with stand variables

The parameters of the mixture distribution were modelled directly with stand variables and percentiles. This is the first attempt of modelling the parameters of mixture distribution with stand variable as much as we know. Stand variables such as stand age, stand density, basal area per ha, quadratic mean diameter, dominant height, variance of the diameter distribution and percentiles were used. The Seemingly Unrelated Regression (SUR) technique was used in this study. Prior to the application of SUR, the candidate models for each response variable (parameters and percentiles) were first identified. Thereafter, the models were simultaneously fitted using SUR. This system of equation is called disaggregation system (Gómez-García et al. 2014). The model structure for the system of equation in matrix form is given by:

$$
\begin{aligned}
y_1 &= X_1 b_1 + \varepsilon 1 \\
&\vdots \\
y_2 &= X_2 b_2 + \varepsilon_2 \qquad \text{Eq. (12)} \\
&\vdots \\
y_n &= X_n b_n + \varepsilon_n
\end{aligned}
$$

Where n=number of component model in the system of equation, y=response variable (the parameters of the Weibull mixture and percentiles); x=predictor variables (e.g., stand age, stand density, basal area per ha, quadratic mean diameter etc.); ε=error term which is assumed to be normal and independently distributed with zero mean and constant variance that is, $\varepsilon \sim NID(0, \sigma^2)$. Prior to model development, 2 plots were set aside because there were few

Table 2. Kolmogorov-Smirnov (Dn) and negative loglikelihood (-ΛΛ) values of the fitted distributions to diameter data

Plot	Kolmogorov-Smirnov (Dn)					Negative loglikelihood (-ΛΛ)				
	GW	SB	LL	GB	FM	GW	SB	LL	GB	FM
1	0.0881	0.0862	0.0819	0.0901	0.0695	156.113	156.019	155.848	156.11	155.756
2	0.0574	0.0577	0.0577	0.0578	0.0640	166.791	166.828	166.485	166.824	168.290
3	nf	0.1739	0.0949	nf	0.0971	nf	125.071	124.301	nf	125.030
4	0.0986	0.1019	0.0934	0.105	0.0784	138.972	138.971	139.052	139.141	136.013
5	0.1221	0.1292	0.1240	0.1267	0.1004	154.243	154.458	154.242	154.358	156.128
6	0.0862	0.0918	0.0850	0.0998	0.0670	204.287	204.464	203.944	204.667	202.052
7	0.0714	0.3585	nf	nf	0.0817	155.156	154.632	nf	nf	156.903
8	0.0979	0.1063	0.0846	0.1232	0.1106	150.257	150.526	149.990	151.111	149.717
9	nf	0.0774	0.0744	0.0823	0.0813	nf	164.447	163.940	163.436	169.905
10	0.0564	0.7200	0.0596	0.0662	0.0455	133.482	133.455	132.559	133.059	134.279
11	0.0959	0.0974	0.0932	0.1002	0.1145	130.040	130.106	130.116	130.335	133.224
12	0.0627	0.0616	0.0625	0.0664	0.0805	191.418	191.455	191.134	191.481	194.725
13	nf	0.1489	nf	0.0826	0.0710	nf	139.549	nf	139.074	142.068
14	0.0600	0.0622	0.0660	0.0628	0.0611	331.134	331.905	331.504	331.696	333.309
15	nf	0.1125	nf	nf	0.0870	nf	138.808	nf	nf	141.686
16	nf	0.3600	nf	nf	0.1052	nf	147.500	nf	nf	150.754
17	0.0912	0.0915	0.0938	0.0935	0.0847	240.850	240.946	240.194	240.909	240.329
18	nf	0.0660	0.0781	0.0904	0.0567	nf	139.658	139.031	138.848	142.797
19	0.1110	0.1035	0.1001	0.1034	0.1161	148.153	148.458	148.006	148.385	145.632
20	0.0957	0.1200	0.1007	0.0989	0.0721	150.961	150.824	150.592	150.719	149.353
21	nf	0.0834	0.0804	nf	0.0949	nf	151.518	151.071	nf	153.620
22	nf	0.6275	nf	nf	0.0920	nf	129.324	nf	nf	134.992
23	0.0631	0.0706	0.0668	0.0684	0.0775	177.982	178.523	178.302	178.286	180.998
24	nf	0.0927	0.0893	0.0868	0.0816	nf	172.044	171.204	171.569	176.446
25	0.0976	0.1282	0.1078	0.1109	0.0881	111.516	111.452	110.820	111.203	111.914
26	0.0647	0.0742	0.0760	0.074	0.0651	162.117	162.179	161.498	162.039	165.172
27	nf	0.5682	nf	nf	0.1054	nf	120.104	nf	nf	120.668
28	nf	0.1698	nf	nf	0.1886	nf	160.374	nf	nf	162.711
29	nf	0.2759	nf	nf	0.0888	nf	164.980	nf	nf	168.725
30	nf	0.3810	nf	nf	0.0559	nf	173.825	nf	nf	179.964
31	nf	0.4038	0.0607	nf	0.1509	nf	135.626	135.135	nf	138.676
32	0.0868	0.2222	nf	nf	0.1073	127.950	127.360	nf	nf	129.310
33	nf	0.4800	0.0601	nf	0.0600	nf	127.958	127.274	nf	131.127
34	nf	0.1387	nf	nf	0.2207	nf	132.624	nf	nf	134.302
35	0.1054	0.1008	0.1003	0.1032	0.0787	172.084	172.195	171.769	172.14	171.569

nf, no fit.

data points (N=35 plots). The two plots were used to evaluate the models in the system.

Richards height-diameter model was fitted to each plot data, this was used to estimate the mean height per diameter class wherein the stand volume was derived. The Richards function (Richards 1959) is expressed as:

$$H = bh + a(1 - e^{-bD})^{c} \qquad \text{Eq. (13)}$$

Where H=height; D=dbh; bh=breast height which was taken at 1.3 m above the ground, a and b=model parameters. The choice of Richards function stemmed from the fact that this function has the sigmoidal properties of monotonic, inflection point and asymptote (Lei and

Table 3. Kolmogorov-Smirnov (Dn) and negative loglikelihood values of the fitted distributions to height data

Plot	Kolmogorov-Smirnov (Dn)					Negative loglikelihood (-ΛΛ)				
	GW	SB	LL	GB	FM	GW	SB	LL	GB	FM
1	0.0717	0.0621	0.0747	0.069	0.0513	111.254	111.434	110.634	111.256	110.916
2	0.0738	0.0754	0.0674	0.0774	0.0673	135.011	135.118	134.867	135.175	133.663
3	0.0582	0.3478	0.0702	0.0788	0.0543	110.524	110.489	109.914	110.342	110.633
4	0.0630	nf	0.0620	0.1051	0.0576	106.143	104.724	104.363	105.005	105.073
5	0.0625	0.0607	0.0596	0.0601	0.0712	115.777	115.775	115.525	115.771	115.642
6	0.0512	0.1622	0.0550	nf	0.0446	185.415	185.000	184.598	nf	185.291
7	0.0931	0.6415	0.0762	0.0856	0.0529	137.313	136.590	136.222	136.475	135.409
8	0.0769	1.0000	0.0739	0.0745	0.0576	123.635	124.684	123.229	123.607	123.698
9	0.1008	nf	0.0915	0.0951	0.0583	139.211	139.193	137.985	138.418	138.176
10	0.0527	0.6000	0.0528	0.0569	0.0462	126.459	126.306	125.771	126.112	126.825
11	0.0627	0.0667	0.0605	0.0586	0.0686	96.164	96.345	96.027	96.353	96.428
12	0.0695	0.7619	0.0656	0.0868	0.0609	163.795	162.837	162.183	163.441	163.406
13	0.0781	0.5957	nf	nf	0.0645	130.031	129.666	nf	nf	131.100
14	0.0471	0.1192	0.0350	nf	0.0431	303.640	302.161	301.784	nf	302.779
15	0.0748	0.0872	0.0803	0.0859	0.0744	121.823	122.529	122.020	122.094	121.789
16	0.2641	0.6200	nf	nf	0.1298	132.053	132.426	nf	nf	132.677
17	0.0550	0.0529	0.0508	0.0583	0.0421	191.578	192.023	191.956	192.143	191.193
18	0.0474	0.3674	0.0505	0.0536	0.1513	110.217	110.159	109.879	110.115	114.090
19	0.0925	0.0710	0.0771	0.0875	0.0594	117.242	118.064	117.700	117.928	118.196
20	0.0864	0.7600	0.0799	0.0875	0.0644	117.588	117.187	116.593	116.833	118.014
21	0.0624	0.5192	0.0693	0.0653	0.0837	134.499	134.779	134.682	134.676	136.937
22	0.1041	0.1013	0.0971	nf	0.0928	140.721	140.588	140.595	nf	138.424
23	nf	0.1967	0.0727	0.0688	0.0949	nf	164.904	164.572	164.575	168.735
24	0.0613	0.0625	0.0612	0.0615	0.0715	133.523	133.658	133.547	133.609	135.078
25	0.0832	0.9231	0.0813	nf	0.0613	89.383	89.189	88.486	nf	90.165
26	0.1194	0.8113	nf	nf	0.1330	133.736	131.515	nf	nf	130.631
27	nf	0.2272	0.0730	0.0739	0.0801	nf	98.043	97.383	97.739	101.217
28	0.0865	0.5849	nf	nf	0.0834	137.475	137.438	nf	nf	140.730
29	nf	0.3965	0.1048	0.1187	0.1852	nf	148.649	147.930	147.738	158.054
30	0.0692	0.8571	0.0590	0.0569	0.1449	154.970	154.870	154.517	154.651	157.266
31	0.0524	0.9423	0.0590	0.0649	0.0497	106.570	106.261	105.621	105.842	106.829
32	0.0517	0.6667	0.0577	0.0583	0.0562	109.960	110.202	109.740	109.972	111.127
33	0.0573	nf	0.0608	0.0579	0.0565	103.140	nf	102.235	102.574	104.314
34	nf	0.9111	nf	nf	0.1329	nf	105.781	nf	nf	110.505
35	0.0560	0.3167	0.0540	0.0594	0.0539	149.717	150.176	149.684	150.029	151.067

nf, no fit.

Parresol 2001). Several researchers have used the Richards function to model height-diameter relationship including Mehtätalo et al. (2015), Özçelik et al. (2018) and Sharma et al. (2018).

Result and Discussion

Fits of the distribution

The Kolmogorov-Smirnov (Dn) and negative loglikelihood (-ΛΛ) values of generalized Weibull (GW), Johnson's S_B (SB), Logit-Logistic (LL), generalized beta (GB) and finite mixture (FM) for fitting diameter and height distributions are presented in Tables 2 and 3. The results showed that FM had the smallest Dn and -ΛΛ values in some of the plots for diameter and height distributions. For example, FM had the smallest Dn value in 20 and 24 out 35 plots for the diameter and height distributions, respectively. The performance of LL, GW and GB distributions was equally good and relatively the same. The SB distribution was the least performed distribution among others. Its indices were generally large. The flexibilities of the four 4-parameter distributions have been demonstrated in Wang and Rennolls (2005). They reported that the LL covers widest area in the skewness and kurtosis plane; the coverage of GB and SB are relatively the same. Distributions with such flexibility should be able to describe the wide varieties of forest stand structure.

Furthermore, the 4-parameter, especially, GW, LL and GB distributions had no result in some of the plots for both diameter and height. The number of plots in which no fit was recorded for GW, LL and GB were 16, 11 and 14, respectively for diameter distribution. In the case of height distribution, no fit was recorded for GW, SB, LL and GB in 4, 1, 5 and 9 plots, respectively. Contrary to the 4-parameter distributions, the FM distribution had reasonable fit to the data from the 35 plots. The plots for which the 4-parameter distributions had no fit were either bimodal or heavily skewed. These irregularities can be associated with natural or anthropogenic activities such as bush burning, thinning, illegal exploitation and so on (Tsogt and Lin 2012; Lin et al. 2016; Ogana 2018). Studies have shown that such data are best fitted with finite mixture distribution (e.g. Zhang et al. 2001; Liu et al. 2002; Zhang and Liu 2006; Jaworski and Podlaski 2012; Tsogt and Lin 2012; Liu et al. 2014; Ogana 2018). The FM distribution used in this study did not only provide reasonable fits to these irregular plots data but also performed best in some of the plots that seems to have regular normal shape (Appendices A

Table 4. Developed system of equations and the fit indices

Model	R^2	RMSE	Eq. no
$\bar{d}=-0.079+0.983d_g-4.728A^{-1}$	0.994	0.180	(14)
$D_5=-4.542+1.028d_g-0.135S_d^2$	0.857	0.885	(15)
$D_{25}=-3.153+1.098d_g-0.129S_d^2$	0.996	0.434	(16)
$D_{50}=-2.201+1.093d_g-0.049S_d^2+0.426\dfrac{H_d}{A}$	0.926	0.666	(17)
$D_{70}=1.699+0.996d_g$	0.872	0.842	(18)
$D_{75}=3.046+0.964d_g$	0.874	0.817	(19)
$H_{75}=-0.218+0.827H_d+0.077A$	0.881	0.672	(20)
$D_{95}=1.619+1.107d_g-0.175S_d^2$	0.901	1.030	(21)
$\alpha_1=9.439-0.693H_{75}+26.281\dfrac{D_5}{D_{95}}$	0.338	2.655	(22)
$\beta_1=27.406+1.321P_{25}+1.308\dfrac{1}{In\left(\dfrac{P_{25}}{d_g}\right)}-20.451\sqrt{\dfrac{D_{25}}{D_{75}}}$	0.915	0.961	(23)
$\alpha_2=15.193-0.124S_d^2-106040ShI$	0.434	1.237	(24)
$\beta_2=4.764+1.704d_g-0.614D_{50}-4.928\dfrac{D_5}{D_{95}}$	0.960	0.487	(25)

and B). The teak plantation has been thinned and encroached by illegal loggers (Chukwu and Osho 2018); hence the reason for the irregular shape in some of the plots.

From the foregoing discussion, it is obvious that 4-parameter distribution cannot describe all possible forest stand structure. Therefore, we conclude that whenever complex distribution is to be used for tree size distribution modelling, a simple finite mixture should be considered such as the simple two components 2-parmaeter Weibull mixture distribution with equal mixing proportion.

Modeling the diameter distribution with FM

Twelve models were developed in the system of equation – one for predicting mean diameter (equation 14), seven for predicting percentiles (equation 15 to 21) and four for predicting the parameters of the FM (equation 22 to 25) as shown in Table 4. In this system, the best prediction models for α_1, β_1, α_2 and β_2 were those involving dominant height (H_d), quadratic mean (d_g), shape index (ShI), variance (S_d^2), 5%, 25%, 75%, 95% of the diameter and 75% of the height distributions. The percentiles were predicted from dominant height, stand age (A), quadratic mean diameter and variance. The variance was derived as the difference between square quadratic mean and square arithmetic mean diameters (*i.e.*, $d_g^2 - \overline{d^2}$). The shape index was derived from basal per ha, number of tree per ha and the basal area of central diameter. The basal area of central diameter was taken as the 70th percentile of the diameter distribution (van Laar and Akça 2007). Thus, the exogeneous variables (instrumental variables) were stand age, dominant height, basal area and number of trees per ha. The coefficient of determination (R^2) and root mean square error (RMSE) of the models in the system ranged from 0.337-0.994 and 0.180-2.655, respectively. The standard errors of the parameters of the models were relatively low. The variance inflation factor (VIF) was generally low; as such, there was little or no autocorrelation in the independent variables.

The system of equation was applied to predict the diameter distribution of two independent plots (1 and 35) *i.e.*, plots not included in developing the system. The results obtained were comparable to the observed diameter distributions of the stand. The estimated number of trees per ha and volume per ha by diameter class for the two plots are presented in Figs. 1, 2. The mean height per diameter class used in computing volume was obtained from the fitted Chapman-Richards function (displayed on the graph). The two plots were not rejected by the Kolmogorov-Smirnov (Dn) test at 5% level ($p > 0.05$).

This system of equation can be used to predict the future yield of the forest stand by first projecting the exogenous variables and then other variables can be derived. This is the first attempt of relating the parameters of FM dis-

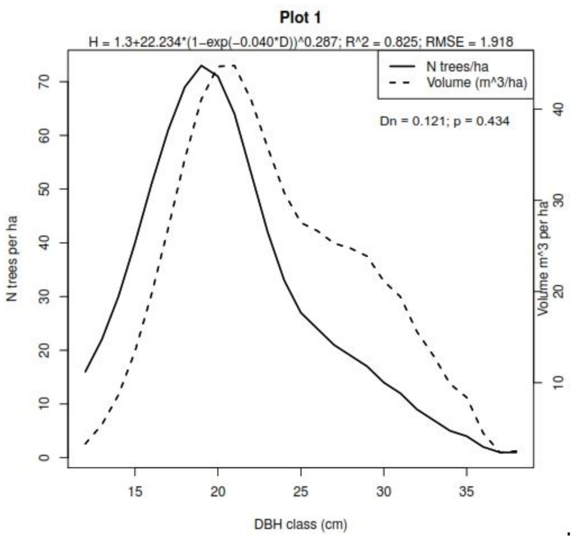

Fig. 1. Number of trees and volume per ha against diameter derived from the system for plot 1.

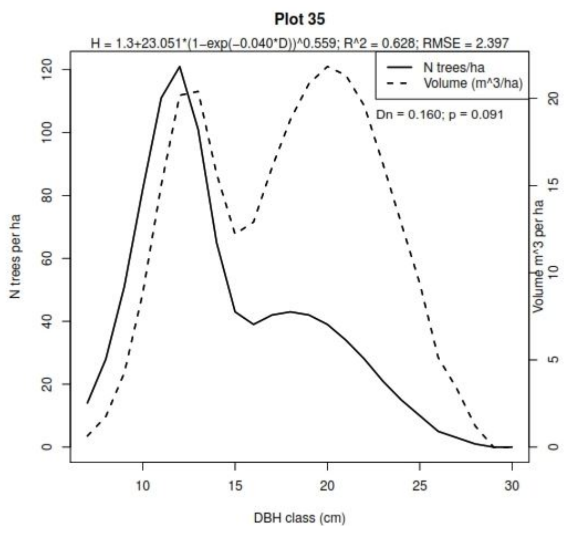

Fig. 2. Number of trees and volume per ha against diameter derived from the system for plot 35.

tribution with stand variables as much as we know. Most published literatures on yield system have been based on single component distribution of either Weibull, beta or Johnson's S_B distributions etc. For example, Stankova and Zlatanov (2010) used three methods including percentile-based projection and a single component 2-parameter Weibull to model the diameter distribution of *Pinus nigra* Arn wherein good results were reported. Gómez-García et al. (2014) also developed a dynamic volume and biomass growth system of equation for birch stand. Similar observation was reported in Cao (2004), Poudel (2011) and Poudel and Cao (2013) where the authors used SUR to develop system of equations. They incorporated a single component Weibull distribution in the system. Furthermore, Gorgoso-Varela et al. (2008) developed a compatible system in which the two shape parameters of the beta distribution were derived from the first and second moments of the diameter distribution i.e., average diameter and variance, respectively. The same procedure was used by Parresol (2003), Fonseca et al. (2009) and Mateus and Tomé (2011) for the Johnson's S_B distribution. The performance of these distributions especially the four-parameter functions cannot be overemphasized. However, these functions cannot uniquely describe the diameter distribution of heavily skewed or multimodal peaks. Application of single component four-parameter distribution will result to poor fit or oversimplification of the tree size distribution (Tsogt and Lin 2012; Ogana 2018). In consequence, poor estimation of the forest stand volume. The predictions from FM were reasonable and outperformed other four-parameter distributions.

Conclusion

This study utilised the simplest form of finite mixture distribution − 2 components with equal mixing proportion (0.5) for modelling simple, complex, irregular or bimodal forest stand structure. The mixture model outperformed the four-parameter distribution functions in terms of their fitness to predict diameter and height structures of the disturbed teak stand. Models for the parameters of the simple mixture distribution were developed with SUR technique and the predictions were reasonable and compare well with the observed distribution. This is the first attempt of relating the parameters of mixture distribution directly with stand variable as much as we know. The simplified mixture distribution can be used to model other tree variables e.g. crown diameter, crown surface area, basal area in the case of relaskop sampling, biomass, etc. in so far, the variable can be measured directly from the field. It can also be integrated as component model in stand density management diagram.

References

Ajayi S. 2013. Diameter Distribution for *Gmelina Arborea* (ROXB) Plantations in Ukpon River Forest Reserve, Cross River State, Nigeria. Afrrev Stech Int J Sci Technol 2: 64-82.

Burkhart HE, Tomé M. 2012. Modeling Forest Trees and Stands. (2nd ed) Springer Dordrecht Heidelberg, New York, pp 271.

Cao QV. 2004. Predicting Parameters of a Weibull Function for Modeling Diameter Distribution. For Sci 50: 682-685.

Chukwu O, Osho JSA. 2018. Basal Area-Stump Diameter Models for *Tectona grandis* Linn. F. Stands in Omo Forest Reserve, Nigeria. J For Environ Sci 34: 119-125.

Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL. 1983. Timber management: a quantitative approach. Wiley, New York, NY, pp 333.

De Liocourt F. 1898. De l'aménagement des sapinières. Bull Soc For Franche-Comtéet Belfort 4: 396-409.

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

Gómez-García E, Crecente-Campo F, Tobin B, Hawkins M, Nieuwenhuis M, Diéguez-Aranda U. 2014. A dynamic volume and biomass growth model system for even-aged downy birch stands in south-western Europe. Forestry 87: 165-176.

Gorgoso-Varela JJ, Garcia-Villabrille JD, Rojo-Alboreca A, Gadow Kv, Alvarez-Gonzalez JG. 2016. Comparing Johnson's SBB, Weibull and Logit-Logistic bivariate distributions for modeling tree diameters and heights using copulas. For Syst 25: 1-5.

Gorgoso-Varela JJ, Rojo-Alboreca A, Afif-Khouri E, Barrio-Anta M. 2008. Modelling diameter distributions of birch (*Betula alba* L.) and pedunculate oak (*Quercus robur* L.) stands in Northwest Spain with the beta distribution. Investigación Agraria: Sistemas y Recursos Forestales 17: 271-281.

Gorgoso-Varela JJ, Rojo-Alboreca A. 2014. A comparison of estimation methods for fitting Weibull and Johnson's SB functions to pedunculate oak (*Quercus robur*) and birch (*Betula pubescens*) stands in northwest Spain. For Syst 23: 500-505.

Jaworski A, Podlaski R. 2012. Modelling irregular and multimodal tree diameter distributions by finite mixture models: an approach to stand structure characterization. J For Res 17: 79-88

Jin X, Li F, Jia W, Zhang L. 2013. Modeling and Predicting Bivariate Distributions of Tree Diameter and Height. Scientia Silvae Sinicae 49: 74-82.

Larsen TN, Cao QV. 2006. A diameter distribution model for even-aged beech in Denmark. For Ecol Manage 231: 218-225.

Lei Y, Parresol BR. 2001. Remarks on height-diameter modeling. Research note SRS, 10. Southern Research Station, Asheville, NC.

Li F, Zhang L, Davis CJ. 2002. Modeling the Joint Distribution of Tree Diameters and Heights by Bivariate Generalized Beta Distribution. For Sci 48: 47-58.

Lin C, Tsogt K, Zandraabal T. 2016. A decompositional stand structure analysis for exploring stand dynamics of multiple attributes of a mixed-species forest. For Ecol Manage 378: 111-121.

Liu C, Zhang L, Davis CJ, Solomon DS, Gove JH. 2002. A Finite Mixture Model for Characterizing the Diameter Distributions of Mixed-Species Forest Stands. For Sci 48: 653-661.

Liu F, Li F, Zhang L, Jin X. 2014. Modeling diameter distributions of mixed-species forest stands. Scan J For Res 29: 653-663.

Mateus A, Tomé M. 2011. Modelling the diameter distribution of *Eucalyptus* plantations with Johnson's S B probability density function: parameters recovery from a compatible system of equations to predict stand variables. Ann For Sci 68: 325-335.

Mehtätalo L, de-Miguel S, Gregoire TG. 2015. Modeling height-diameter curves for prediction. Can J For Res 45: 826-837.

Ogana FN, Itam ES, Osho JSA. 2017. Modeling diameter distributions of *Gmelina arborea* plantation in Omo Forest Reserve, Nigeria with Johnson's S_B. J Sustain For 36: 121-133.

Ogana FN, Osho JSA, Gorgoso-Varel JJ. 2018. Application of Extreme Value Distribution for Assigning Optimum Fractions to Distributions with Boundary Parameters: An Eucalyptus Plantations Case Study. Sib J For Sci 4: 39-48.

Ogana FN. 2018. Application of Finite Mixture to Characterise Degraded *Gmelina arborea* Roxb Plantation in Omo Forest Reserve, Nigeria. J For Environ Sci 34: 451-456.

Özçelik R, Cao QV, Trincado G, Göçer N. 2018. Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey. For Ecol Manage 491(420): 240-248.

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). Eur J For Res 126: 563-571.

Parresol BR. 2003. Recovering parameters of Johnson's SB distribution. Research paper SRS, 31. US Dept of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Podlaski R. 2017. Forest modelling: the gamma shape mixture model and simulation of tree diameter distributions. Ann For Sci 74: 29.

Poudel KP, Cao QV. 2013. Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions. For Sci 59: 243-252.

Poudel KP. 2011. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. MS thesis. Louisiana State University and Agricultural and Mechanical College, Louisiana, USA. (in English)

R Core Team. 2017. R: A language and environment for statistical computing. http://www.R-project.org/. Accessed 30 Jun 2017.

Richards FJ. 1959. A Flexible Growth Function for Empirical Use. J Exp Bot 10: 290-301.

Sharma RP, Vacek Z, Vacek S, Kučera M. 2018. Modelling individual tree height-diameter relationships for multi-layered and multi-species forests in central Europe. Trees 33: 103-119.

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. Eur J For Res 129: 1169-1179.

Tsogt K, Lin C. 2012. A flexible modeling of irregular diameter structure for the volume estimation of forest stands. J For Res 19: 1-11.

van Laar A, Akça A. 2007. Forest mensuration. Springer, Dordrecht, Netherlands, pp 389.

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

Wang M. 2005. Distributional modelling in forestry and remote sensing. PhD thesis. University of Greenwich, Greenwich, UK. (in English)

Weibull W. 1951. A Statistical Distribution Function of Wide Applicability. J Appl Mech 18: 293-297.

Zasada M, Cieszewski CJ. 2005. A finite mixture distribution approach for characterizing tree diameter distributions by natural social class in pure even-aged Scots pine stands in Poland. For Ecol Manage 204: 145-158.

Zhang L, Gove JH, Liu C, Leak WB. 2001. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. Can J For Res 31: 1654-1659.

Zhang L, Liu C. 2006. Fitting irregular diameter distributions of forest stands by Weibull, modified Weibull, and mixture Weibull models.

Zhang L, Packard KC, Liu C. 2003. A comparison of estimation methods for fitting Weibull and Johnson's S_B distributions to mixed spruce-fir stands in northeastern North America. Can J For Res 33: 1340-1347.