RESEARCH ARTICLE

Evaluation of direct and indirect methods for modelling the joint distribution of tree diameter and height data with the bivariate Johnson's SBB function to forest stands

José Javier Gorgoso-Varela[1], Friday Nwabueze Ogana[2] and Rafael Alonso-Ponce[1,3]

[1]föra forest technologies sll. Campus Duques de Soria. 42004 Soria. Spain. [2]Department of Social and Environmental Forestry, University of Ibadan, Ibadan, Nigeria. [3]Sustainable Forest Management Research Institute University of Valladolid-INIA, Campus Duques de Soria. 42004 Soria. Spain.

Abstract

Aim of study: In this study, both the direct and indirect methods by conditional maximum likelihood (CML) and moments for fitting Johnson's S_{BB} were evaluated. To date, Johnson's S_{BB} has been fitted by either indirect (two-stage) method using well-known procedures for the marginal diameter and heights, or direct methods, where all parameters are estimated at once. Application of bivariate Johnson's S_{BB} for predicting height and improving volume estimation requires a suitable fitting method.

Area of study: E. globulus, P. pinaster and *P. radiata* stands in northwest Spain.

Material and methods: The data set comprised of 308, 184 and 96 permanent sample plots (PSPs) from the aforementioned species. The suitability of the method was evaluated based on height and volume prediction. Indices including coefficient of determination (R^2), root mean square Error (RMSE), model efficiency (MEF), Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC) were used to assess the model predictions. Significant difference between observed and predicted tree height and volumes were tested using paired sample t-test at 5% level for each plot by species.

Main results: The indirect method by CML was the most suitable method for height and volume prediction in the three species. The R^2 and RMSE for height prediction ranged from 0.994 – 0.820 and 1.454 – 1.676, respectively. The percentage of plot in which the observed and predicted heights were significant was 0.32%. The direct method was the least performed method especially for height prediction in *E. globulus.*

Research highlights: The indirect (two-stage) method, especially by conditional maximum likelihood, was the most suitable method for the bivariate Johnson's S_{BB} distribution.

Additional keywords: conditional maximum likelihood; moments; two-stage method; direct method; tree volume.

Introduction

Researchers realize that volume, the primary variable that forest managers are interested in, is heavily dependent on both tree diameter and height. A traditional practice is to fit a marginal distribution to the diameter frequency data and then use an empirical height-diameter relationship to estimate the average height per diameter class and hence the volume (Clutter & Allison, 1974). Although this approach is satisfactory

in many situations, it is often not appropriate because the approach tends to ignore the natural relationships between tree diameters and heights by treating them separately (Schreuder & Hafley, 1977; Tewari & von Gadow, 1999).

The traditional method does not quantify the distribution of heights for a given diameter and one approach for modelling the conditional height distribution for the different diameters is to use the height residuals (Gaffrey, 1996). However, it is very seldom that the height residuals are homoscedastic and normally distributed. In most forest stands the variance about the diameter-height regression is not homogeneous (Zucchini et al., 2001).

One of the most important elements of forest structure is the relationship between tree diameters and heights because information about size-class distributions of the trees within a forest stand is important for estimating product yields (Zucchini et al., 2001). The size-class distribution influences the growth potential and hence the current and future economic value of a forest stand (Knoebel & Burkhart, 1991). On the other hand, the social status of a tree, which reflects its further development in the aspects of growth and mortality, depends not only on its relative diameter, but also on its relative height in a stand (Siipilehto, 2000). Furthermore, knowledge of the height variation, both between and within diameter classes, improves the chances of successfully imitating different types of thinnings (Hafley & Buford, 1985).

For many years, the bivariate extension of the S_B distribution, the S_{BB} (Johnson, 1949a and 1949b), has been the most commonly bivariate distribution used for modeling bivariate tree diameter-height frequency data (e.g. Schreuder & Hafley, 1977; Hafley & Buford, 1985; Knoebel & Burkhart, 1991; Siipilehto, 1996; Tewari & von Gadow, 1997; Tewari & von Gadow, 1999; Zucchini et al., 2001; Li et al., 2002; Wang & Rennolls, 2007, Gorgoso-Varela et al., 2016; Ogana et al., 2018; Ogana, 2018a). The Johnson's S_{BB} is developed by applying a four-parameter logistic transformation to each of the component variables of a standard bivariate normal distribution (Johnson, 1949b; Wang, 2005; Gorgoso-Varela et al., 2016). The fitting of the joint distribution of diameters and heights is called bivariate distribution modeling (Ogana et al., 2018). Bivariate distribution modelling allows for the generation of bivariate frequencies of tree diameter and height (Tewari & von Gadow, 1997).

The parameters of the Johnson's S_{BB} distribution have been estimated with both the indirect and direct methods. The indirect method fits the bivariate Johnson's S_{BB} distribution by two-stage where the marginals are first fitted separately for the diameter and height; the estimates are then used to compute the parameter of association. Examples of the indirect method that have been reported in forestry literature include Conditional Maximum Likelihood (CML), moments, mode and Knoebel and Burkhart methods (Hafley & Buford, 1985; Tewari & von Gadow, 1997; Knoebel & Burkhart, 1991; Siipilehto, 1996; Tewari & von Gadow, 1999; Castedo-Dorado et al., 2001; Li et al., 2002; Ogana, 2018a). Of these indirect methods, CML and moments have been reported as the best methods (Ogana, 2018a). In the direct method, all the parameters in the bivariate distribution are estimated simultaneously (Wang, 2005; Mønness, 2015; Gorgoso-Varela et al., 2016; Ogana et al., 2018). However, these two methods (i.e., direct and indirect) of estimating the parameters of Johnson's S_{BB} distribution have never been compared using the same sample. The method used to calibrate the distribution matters in order to evaluate the accuracy of diameter and height predictions (Ogana, 2018a). Thus, the objective of the study was to compare the indirect fitting method for the Johnson's S_B based on two-stage method by the Conditional Maximum Likelihood and moments methods with the direct method where the parameters are estimated simultaneously.

Material and methods

Data set

The data used for this study were obtained from three temperate species – Tasmanian blue gum (Eucalyptus globulus Labill) and two species of pine – Maritime pine (Pinus pinaster Ait) and Monterrey pine (Pinus radiata D. Don) in Spain. E. globulus stands occupy 320,774 ha in Galicia. The pure stands of P. pinaster cover 217,281 ha and 22,523 ha in the regions of Galicia and Asturias, respectively. The plantations of P. radiata occupy 96,177 ha and 25,385 ha in Galicia and Asturias, respectively (MMAMRM, 2011). A total of 308 permanent sample plots (PSPs) from E. globulus, 184 PSPs from P. pinaster stands and 96 PSPs from P. radiata stands were used for this study. The plot sizes ranged from 375 to 900 m²; to achieve a minimum of 30 trees per plot. Diameter at breast height (Dbh at 1.3 m above the ground) and total height were measured with calliper and hypsometer to a precision of the nearest 0.1 cm and 0.1 m, respectively. A total of 16382, 17845 and 12722 trees were measured from E. globulus, P. pinaster and P. radiata, respectively. The descriptive statistics of the inventory data are presented in Table 1.

Species	Stand variable	Statistics					
		Mean	Max	Min	SD	Skw	Kurt
E. globulus	d (cm)	12.9	69.8	3.8	5.79	0.98	2.46
(N plots = 308; N trees = 17,582)	h (m)	15.2	42.6	0.9	5.79	0.60	0.56
P. pinaster	d (cm)	17.5	63.2	5.1	6.84	1.32	3.29
(N plots = 184; N trees = 17,960)	h (m)	11.5	36.0	3.6	3.73	1.02	2.17
P. radiata	d (cm)	16.5	47.1	5.1	6.05	0.53	0.21
(N plots = 96; N trees = 12,895)	h (m)	14.7	31.7	3.8	3.85	0.29	0.15

Max = maximum, min = minimum, SD = standard deviation, Skw = skewness, Kurt = kurtosis, d = tree diameter at 1.30 m, h = total height.

The Johnson's Univariate (S_B) and Bivariate Distribution (S_{BB})

The univariate Johnson's S_B distribution (Johnson, 1949a) is expressed as:

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

Where:

$\xi < x < \xi + \lambda$, $-\infty < \xi < +\infty$, $-\infty < \gamma < +\infty$, $\lambda > t0$, and $\delta > 0$. ξ shape parameters (asymmetry and kurtosis parameters, respectively). The Johnson's S_{BB} (Johnson, 1949b) is the extension of the univariate Johnson's S_B distribution; given by:

$$f(Z_d, Z_h, \rho) = \frac{1}{(2\pi\sqrt{1-\rho^2})} \exp$$
$$\exp\left[-\frac{1}{2}(1-\rho^2)^{-1}(Z_d^2 - 2\rho Z_d Z_h + Z_h^2)\right] \qquad \text{Eq. (2)}$$

Where:

$$\text{Where: } \rho = \sum_{i=1}^{n} Z_{di} Z_{hi}/n \qquad \text{Eq. (3)}$$

$$Z_d = \gamma_d + \delta_d \ln\left(\frac{d-\xi_d}{\xi_d+\lambda_d-d}\right) \qquad \text{Eq. (4)}$$

$$Z_h = \gamma_h + \delta_h \ln\left(\frac{h-\xi_h}{\xi_h+\lambda_h-h}\right) \qquad \text{Eq. (5)}$$

Where: ρ is the correlation coefficient between Z_d and Z_h, the subscripts *d* and *h* in equations represent diameter and height, respectively.

Fitting Methods

Indirect method: this method fits the bivariate Johnson's S_{BB} distribution by a two-stage approach

where the marginals are first fitted separately for the diameter and height. The estimates from the first stage are then used to compute the correlation parameter (as shown in Eq 3). Generally, fitting Johnson's S_{BB} distribution requires the location (ξ) and scale (λ) to be predetermined (related to minimum and range of the tree variable) while δ and γ can be analytically deduced through different approaches (Schreuder & Hafley, 1977). The method of CML and moments were the indirect methods considered in this study as recommended by Ogana (2018a).

Conditional Maximum Likelihood (CML): The values of the parameters were obtained with these expressions:

$$\delta = \frac{1}{s_i} \qquad \text{Eq. (6)}$$

$$\gamma = \frac{-\bar{f_i}}{s_i} \qquad \text{Eq. (7)}$$

$$\bar{f_i} = \sum_{i=1}^{n} \frac{f_i}{n} \qquad \text{Eq. (8)}$$

$$S_i^2 = \frac{\sum_{i=1}^{n}(f_i-\bar{f_i})^2}{n} \qquad \text{Eq. (9)}$$

$$f_i = \ln\left(\frac{x_i-\xi}{\xi+\lambda-x_i}\right) \qquad \text{Eq. (10)}$$

Where: $x_i (i = 1, 2, …, n)$ = tree diameters and heights. The location (ξ) and scale (λ) parameters were set to equal minimum diameter times 0.75 ($0.75 * D_{min}$) and maximum diameter, respectively for the marginal distribution of diameter. These parameters were set at 1.3 and maximum height for the marginal height distribution. The factor 0.75 was adopted because Gorgoso *et al.* (2012) used 0.75 D_{min} for this parameter ξ to achieved good result compared to other constraints in *Pinus pinaster* and *Pinus radiata*. The same species from the same region were used in the study at hand. The CML has been applied to fit the Johnson's S_{BB} by different authors, including Knoebel & Burkhart

(1991); Siipilehto (1996); Tewari & von Gadow (1997); Tewari & von Gadow (1999); Castedo-Dorado *et al.* (2001); Ogana (2018a).

Method of Moments: this method is based on the functional relationship between the first and second moments of diameter and height and the parameters of the S_{BB} distribution. Ogana (2018a) recently used this method to fit the S_{BB} distribution. It is given by:

$$\delta = \frac{\mu(1-\mu)}{Sd(x)} + \frac{Sd(x)}{4}\left[\frac{1}{\mu(1-\mu)} - 8\right] \qquad \text{Eq. (11)}$$

$$\gamma = \delta \ln\left(\frac{1-\mu}{\mu}\right) + \left(\frac{0.5-\mu}{\delta}\right) \qquad \text{Eq. (12)}$$

$$\mu = \frac{\bar{X}-\xi}{\lambda} \qquad \text{Eq. (13)}$$

$$Sd(x) = \frac{\sigma_x}{\lambda} \qquad \text{Eq. (14)}$$

Where: $\bar{X}$ = mean diameters and heights of the plot, σ_x = plot diameter and height standard deviations and $Sd(x)$ = modified standard deviation. The same procedure with CML was used for λ and ξ.

Direct method: in this method, all parameters in the bivariate Johnson's S_{BB} distribution were estimated simultaneously using maximum likelihood estimation. However, the location (ξ) and scale (λ) parameters were predetermined with the same procedure as the indirect method. This method was applied by Wang (2005), Wang *et al.* (2008) and Gorgoso-Varela *et al.* (2016). These authors used the normal copula for this distribution. The joint density of the S_{BB} function derived from the normal copula is given by:

$$h(d,h) = \frac{f(d)g(h)}{\sqrt{1-\rho^2}} e^{-\left[\frac{\rho}{2(1-\rho^2)}\right]\left[\rho(z_d^2+z_h^2)-2\rho z_d z_h\right]} \qquad \text{Eq. (15)}$$

and the likelihood function is expressed as:

$$\ell(\theta|d,h) = \prod_{i=1}^{N} f(d)g(h)\frac{1}{\sqrt{1-\rho^2}} e^{-\left[\frac{\rho}{2(1-\rho^2)}\right]\left[\rho(z_d^2+z_h^2)-2\rho z_d z_h\right]} \qquad \text{Eq. (16)}$$

where $f(d)$ and $g(h)$ are marginal Johnson SB densities for diameter (d) and height (h), respectively defined in equation 1; θ parameters to the estimated and N is the number of observations.

The indirect method (CML and moments) estimations were obtained using SAS/STAT™ software (SAS Institute 2003). The computation of the direct method was carried out using the 'optim' (optimal)

function in R (R Core Team, 2017). The Nelder-Mead simplex algorithm (Nelder & Mead, 1965) was used for the 'optim'.

Model Evaluation

The suitability of the direct and indirect methods for fitting S_{BB} distribution was evaluated by tree height and volume predictions. Prediction of expected tree height given diameter and the estimation of stand volume are the main applications of bivariate distribution. Individual tree heights were predicted with the S_{BB} median regression expressed as:

$$h = \frac{\hat{\lambda}_h \theta}{\left[\left(\frac{\hat{\xi}_d+\hat{\lambda}_d-d}{d-\hat{\xi}_d}\right)^\phi + \theta\right]} + \hat{\xi}_h \qquad \text{Eq. (17)}$$

Where: $\theta = \exp\left[\frac{\hat{\rho}\hat{\gamma}_d-\hat{\gamma}_h}{\hat{\delta}_h}\right]$; $\phi = \hat{\rho}\hat{\delta}_d \Big/ \hat{\delta}_h$ and $\hat{\xi}_d, \hat{\lambda}_d, \hat{\delta}_d, \hat{\gamma}_d$, and $\hat{\xi}_h, \hat{\lambda}_h, \hat{\delta}_h, \hat{\gamma}_h$ are the estimated parameters from the direct and indirect methods. h = total tree height (m) and d = tree diameter at 1.30 m (cm). The ϕ parameter determines the shape of the regression curve while ρ influences the slope. The relationship can only be linear if $\hat{\rho}\hat{\delta}_d = \hat{\delta}_h$ and $\hat{\rho}\hat{\gamma}_d = \hat{\gamma}_h$.

Individual tree volume equation developed by García-Villabrille (2015) for *E. globulus* and Diéguez-Aranda *et al.* (2009) for *P. pinaster* and *P. radiata* were used to obtain observed and predicted tree volume. The procedure of Li *et al.* (2002) and Ogana (2018a) was used in this study. Observed tree volume was computed by substituting the observed diameter and height into the individual tree volume equations. The predicted individual tree volumes were obtained from the observed diameter and predicted tree height by S_{BB} height-diameter median regression using direct and indirect methods:

E. globulus (García-Villabrille, 2015): Eq. (18)
$v = 2.993 \times 10^{-5} \, d^{1.973} \, h^{1.059}$

P. pinaster (Diéguez-Aranda *et al.*, 2009): Eq. (19)
$v = 3.974 \times 10^{-5} \, d^{1.876} \, h^{1.079}$

P. radiata (Diéguez-Aranda *et al.*, 2009): Eq. (20)
$v = 4.851 \times 10^{-5} \, d^{1.883} \, h^{1.004}$

Where: v = total volume of the stem with bark (m³); d = tree diameter (cm) and h = tree height (m). The root mean square error reported by the authors for equation 18, 19 and 20 were 0.0116, 0.0156 and 0.0131, respectively.

Different fit indices including coefficient of determination (R^2), root mean square error (RMSE), model efficiency (MEF) (Mayer & Butler, 1993), Bayesian Information Criterion (BIC) and Hannan-Quinn Criterion (HQC) were used to assess the adequacy of the methods for height and volume predictions. In addition, significant difference between observed and predicted tree heights and volumes were tested using paired sample t-test at 5% level for each plot by species. Plot was rejected if significant difference occurs between observed height/volume and predicted height/volume for that plot at $p<0.05$. Pairwise comparison was done with t-test because outlier was not detected in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \qquad \text{Eq. (21)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p}} \qquad \text{Eq. (22)}$$

$$MEF = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \qquad \text{Eq. (23)}$$

$$BIC = n \ln\left(\frac{RSS}{n}\right) + p \ln n \qquad \text{Eq. (24)}$$

$$HQC = n \ln\left(\frac{RSS}{n}\right) + p \ln \ln n \qquad \text{Eq. (25)}$$

Where: RSS = residual sum of square, n = sample size, p = number of parameters; $\bar{Y}_i$ = average tree height or volume; Y_i is the observed value and $\hat{Y}_i$ is the theoretical value predicted by the model.

Result

The descriptive statistics i.e., the mean, maximum, minimum and standard deviation of the estimated parameters of the bivariate Johnson's S_{BB} distribution are presented in Table 2. The location parameter of the marginal height distribution was set to 1.3 for both direct and indirect methods across the three species. The shape, scale and rho (ρ) parameters of the indirect methods (CML and moments) show little variations. In fact, the standard deviations of the parameters' values were the same at least up to 1 decimal place. This indicates that both CML and moments are closely related. However, estimates from the direct method were different. The standard deviations of the estimated parameters are smaller than that of CML and moments of the indirect method.

The assessment of the suitability of indirect (CML and moments) and direct methods for tree height prediction showed that the indirect method, especially the CML method, had the highest R^2 and the lowest RMSE, MEF, BIC and HQC in *E. globulus*, *P. pinaster* and *P. radiata* (shown in Table 3). The R^2, RMSE, MEF, BIC and HQC of CML ranged from 0.994 – 0.820, 1.454 – 1.676, 0.055 – 0.179, 4559 – 13400 and 4492 – 13335, respectively. However, the direct method had relatively lower R^2 and larger RMSE, MEF, BIC and HQC in *E. globulus*, *P. pinaster* and *P. radiata* relative to the indirect methods. The number of plot in which the observed and predicted values were significant at 5% level (t-test) was 1 out of 308 plots (equivalent to 0.32%) in *E. globulus* for CML. Eight (2.59%) and 14 (4.45%) plots were significant for moments and the direct methods, respectively in *E. globulus*. There was no significant difference between the observed and predicted tree height by CML and moments across the 184 and 96 plots in *P. pinaster* and *P. radiata*, respectively. But 4 and 5 plots were significant with the direct method in *P. pinaster* and *P. radiata*, respectively at the specified level.

In the case of tree volume prediction, indirect (CML and moments) and direct methods performed relatively the same (Table 4). The methods had the same R^2, RMSE and MEF values up to 2 decimal places. The direct method had lowest BIC and HQC in *E. globulus* and *P. radiata*. The number of plot in which the observed and predicted values was significant at 5% level for CML was like the result in tree height prediction (i.e., 1, 0, 0). However, 6, 6, and 3 plots were significant in *E. globulus, P. pinaster* and *P. radiata*, respectively for the direct method.

The mean of the residuals from the height-diameter median regression was computed across DBH classes and plotted accordingly for the three species. The data were group into DBH classes of 5 cm interval and the mean residual prediction for each class by species was assessed. The graphs showed that CML and moments over- and under-estimated the tree height in the larger diameter classes (> 42.5 cm) in *E. globulus* and *P. pinaster* (Fig. 1a to c). Nevertheless, their prediction errors lie within the range of – 0.5 to 0.5, except in the larger DBH classes. Conversely, the direct method did not only over- and under-estimate tree height in the larger diameter classes but also over-estimate tree heights in the lower diameter classes in *E. globulus*. The mean residual height prediction for both direct and indirect method were relatively poor in *P. radiata*.

Furthermore, the graph of residual against predicted tree volume showed that CML and moments and the direct methods occupied the same horizontal band (Fig. 2). Their mean residuals plot lied within the range of -0.6 to 0.6, -0.6 to 0.6 and -0.4 to 0.2 in *E. globulus, P. pinaster* and *P. randiata*, respectively. Such plot is

Table 2. Descriptive statistic of the Johnson's S_{BB} parameters for the three species.

Species	Parameters	Indirect method: CML, $\xi_h = 1.3$				Indirect method: Moments, $\xi_h = 1.3$				Direct method, $\xi_h = 1.3$			
		Mean	Max	Min	SD	Mean	Max	Min	SD	Mean	Max	Min	SD
E. globulus	ξ_d	2.358	8.475	0.060	1.268	2.358	8.475	0.060	1.268	4.355	8.475	3.788	0.749
	λ_d	22.167	69.800	2.600	8.271	22.151	69.800	2.600	8.226	22.402	69.800	6.100	8.067
	δ_d	1.006	2.019	0.555	0.238	1.046	2.027	0.599	0.261	0.934	2.545	0.522	0.280
	γ_d	0.236	1.308	-1.013	0.395	0.223	1.453	-1.193	0.440	0.470	2.503	-0.203	0.355
	ρ	0.899	0.983	0.686	0.051	0.899	0.983	0.686	0.051	0.908	1.000	0.001	0.073
	λ_h	20.851	42.600	3.300	6.857	20.844	42.600	3.300	6.847	21.100	42.600	6.600	6.696
	δ_h	1.237	2.457	0.601	0.319	1.197	2.378	0.555	0.326	1.142	3.076	0.488	0.354
	γ_h	-0.820	1.211	-3.091	0.632	-0.814	1.295	-2.803	0.610	-0.970	-0.103	-3.686	0.476
P. pinaster	ξ_d	5.277	21.038	1.200	2.775	4.990	21.038	1.200	2.784	5.885	21.038	3.788	2.414
	λ_d	28.223	63.150	14.000	7.626	28.918	63.150	14.000	7.428	28.292	63.150	14.000	7.575
	δ_d	1.399	2.102	0.785	0.196	1.492	2.210	0.848	0.229	1.237	2.436	0.532	0.236
	γ_d	0.451	1.944	-0.684	0.475	0.472	1.889	-0.529	0.480	0.491	1.817	-0.331	0.382
	ρ	0.712	0.931	0.039	0.097	0.713	0.931	0.039	0.094	0.703	0.979	0.0001	0.194
	λ_h	15.216	36.000	6.100	4.418	15.421	36.000	6.100	4.355	15.262	36.000	6.100	4.384
	δ_h	2.092	3.183	0.870	0.371	2.176	3.419	0.769	0.420	1.861	3.635	0.634	0.374
	γ_h	-1.491	0.619	-3.376	0.684	-1.440	0.701	-3.301	0.691	-1.369	0.492	-3.023	0.562
P. radiata	ξ_d	3.690	11.738	0.938	1.596	3.690	11.738	0.938	1.596	4.557	11.738	3.788	1.173
	λ_d	30.030	47.050	20.650	4.943	30.030	47.050	20.650	4.943	30.061	47.050	20.650	4.954
	δ_d	1.150	1.828	0.837	0.197	1.182	1.886	0.827	0.224	0.945	1.166	0.673	0.199
	γ_d	0.424	1.786	-0.287	0.287	0.428	1.813	-0.318	0.302	0.451	1.391	-0.099	0.223
	ρ	0.804	0.920	0.618	0.065	0.804	0.920	0.618	0.065	0.838	0.949	0.328	0.104
	λ_h	20.942	31.700	15.200	3.329	20.942	31.700	15.200	3.329	20.963	31.700	15.200	3.334
	δ_h	1.509	2.278	1.003	0.256	1.513	2.298	0.944	0.282	1.262	2.008	0.867	0.237
	γ_h	-0.942	-0.089	-2.562	0.388	-0.954	-0.086	-2.594	0.393	-0.818	-0.233	-2.297	0.309

The subscripts d and h are diameter and height; ρ = rho; ξ, λ, = location and scale parameters; δ, γ = shape parameters; CML = conditional maximum likelihood.

used to assess whether the residual is homoscedastic (i.e. constant variance) or heteroscedastic. The residual of the methods assessed were relatively constant.

Discussion

The direct and indirect methods of fitting Johnson's S_{BB} have been evaluated. The methods follow similar trend across *E. globulus, P. pinaster,* and *P. radiata* with CML being the best for tree height and volume predictions. The number of plots rejected by t-test increased from CML to moments and to the direct method. The percentage of rejections in *E. globulus* was higher than the other species because some plots of coppice stands were considered, in which the minimum diameter considered was less than 5 cm (value of the minimum diameter considered for *Pinus pinaster* and *Pinus radiata* stands). There were also biases in the larger diameter classes (> 42.5 cm) for all methods. Kalbi

et al. (2017) attributed this bias to the small number of trees in the larger diameter class. Few trees were found in the diameter classes > 42 cm across the three species. The Johnson's height-diameter function often perform well if parameter ϕ i.e., the determinant of the regression curve is greater than one (Tewari & von Gadow, 1999). In this study, ϕ had values greater than one in some of the sample plots for CML, moments and direct methods. The model performed relatively well in these plots.

Furthermore, the values of the BIC and HQC difference show that the method of CML fits the diameter-height data better than direct methods for height prediction in *E. globulus, P. pinaster* and *P. radiata.* As a rule of thumb, a minimum ΔAIC/ΔBIC/ΔHQC of ≤ 2 is required for two models to be indistinguishable (Ogana, 2018b; Tewari & Singh, 2018). Though, the measures of the accuracy of the estimates (RMSE) and the relative measures of the model performance (MEF) look quite similar for all methods, the CML seems to be slightly better. Parallel observation was reported in

Table 3. Fit indices and number of plot rejection by t-test at α = 0.05 for height prediction.

Species	Methods	Fit indices					t-test (α = 0.05)
		R²	RMSE	MEF	BIC	HQC	N Plot sig
E. globulus	CML	0.944	1.454	0.055	13264	13197	1
	Moments	0.944	1.455	0.055	13271	13204	8
	Direct	0.926	1.579	0.074	15054	14987	14
P. pinaster	CML	0.911	1.132	0.089	4559	4492	0
	Moments	0.908	1.142	0.091	4863	4795	0
	Direct	0.904	1.145	0.096	4870	4803	4
P. radiata	CML	0.820	1.676	0.179	13400	13335	0
	Moments	0.820	1.678	0.180	13439	13374	0
	Direct	0.808	1.678	0.192	13230	13165	5

RME = root mean square error, MEF = model efficiency; BIC = Bayesian information criterion, HQC = Hannan-Quinn criterion, N Plot sig = number of plot that are significant at 5% level.

Table 4. Fit indices and number of plot rejection by t-test at α = 0.05 for volume prediction.

Species	Methods	Fit indices					t-test (α = 0.05)
		R²	RMSE	MEF	BIC	HQC	N Plot sig
E. globulus	CML	0.990	0.019	0.009	32592	32569	1
	Moments	0.990	0.019	0.009	32975	32953	14
	Direct	0.991	0.019	0.009	29503	29481	6
P. pinaster	CML	0.984	0.029	0.015	40578	48555	0
	Moments	0.983	0.029	0.016	49157	49135	0
	Direct	0.983	0.029	0.017	48769	48747	6
P. radiata	CML	0.981	0.023	0.019	24677	24655	0
	Moments	0.981	0.023	0.019	24882	24860	2
	Direct	0.980	0.023	0.019	24466	24444	3

RMSE = root mean square error, BIC = Bayesian information criterion, HQC = Hannan-Quinn criterion, N Plot sig = number of plots that are significant at 5% level.

Ogana (2018a), who evaluated the performance of CML, moments, Knoebel and Burkhart (KB) and mode methods for height and volume predictions. The author found CML and moment to be the best methods for predicting tree height and volume in *E. camaldulensis* Dehn.

It is obvious from this study that when a complex fitting approach does not outperform another simpler one it is convenient to take the latter. Estimating the parameters of Johnson's S_{BB} by CML or moments is much easier and pose little computational difficulty. The CML and moments estimation procedure can even be carried out on Microsoft excel platform. However, the direct method involves complicated algorithm using maximum likelihood technique. This often requires writing the log-likelihood function and the specification of initial values for the parameters which may not achieve convergence (Wang, 2005). Given the complexity of the likelihood function, its gradient is not tractable and then numerical optimization methods have

to be used (i.e. through the 'optim' function), which may be less efficient than those based on derivatives. The computation procedure gets more complex as the number of parameters in the distribution increases. Bivariate Johnson's S_{BB} has nine parameters which makes fitting somewhat complicated. Often constraints are imposed on the two location and scale parameters of the bivariate distribution to make the estimates more plausible (Wang & Rennolls, 2007). Mønness (2015) used maximum likelihood estimation to fit the Johnson's S_{BB} distribution and compared the result with power-normal and hyperbolic height prediction model. The prediction from power-normal and hyperbolic models were better than bivariate Johnson S_{BB} fitted with maximum likelihood. Other studies that have used maximum likelihood technique to fit the bivariate Johnson's S_{BB} distributions include Wang *et al.* (2008), Gorgoso-Varela *et al.* (2016) and Ogana *et al.* (2018). Satisfactory results were obtained by the authors for
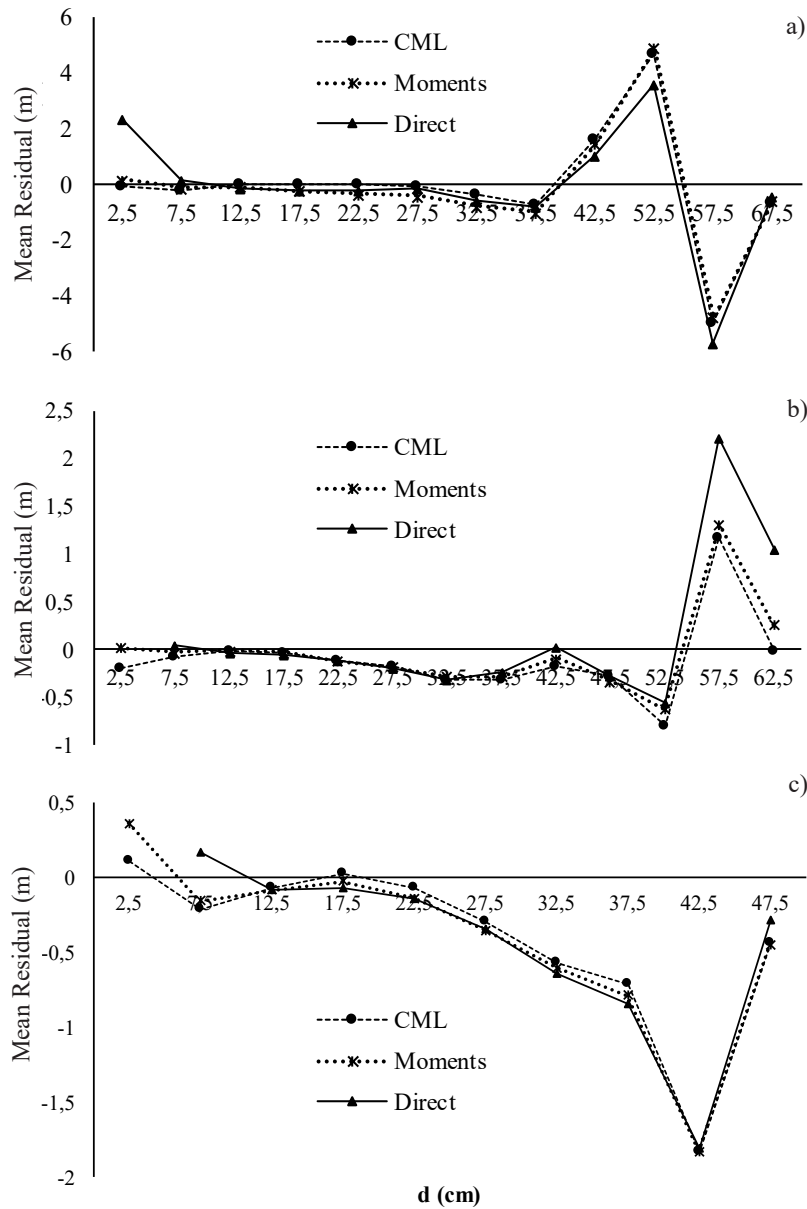
Figure 1. Mean height residual against diameter at 1.30 m of CML, moments and direct methods in (a) *E. globulus*, (b) *P. pinaster* and (c) *P. radiata.*

the Johnson's S_{BB} relative to other bivariate functions evaluated in their study.

The bivariate Johnson's S_{BB} height-diameter median regression remains the most applied function amidst several bivariate distributions in forestry literature. This is because of its flexibility and, more importantly, because its implied relationship between height and diameter is biologically reasonable (Burkhart & Tome, 2012). The closed-form median regression of the Johnson's S_{BB} height-diameter model has also enhanced its continuous application to forestry. This median regression is frequently used for computing percentile lines wherein bounds on height are set. These percentile lines can be used to show how the variation in height

decreases with increasing tree size for a specified diameter (Tewari & von Gadow, 1999). The scatter plots of the Johnson's S_{BB} median regression using indirect method fitted by CML from some sample plots in *E. globulus, P. pinaster* and *P. radiata* are presented in Figures S1 to S3 [suppl.].

Conclusion

Modelling the joint distribution of diameter and height by Johnson S_{BB} remains an important tool for assessing the variation of tree height for a given diameter, detailed stand structure and volume
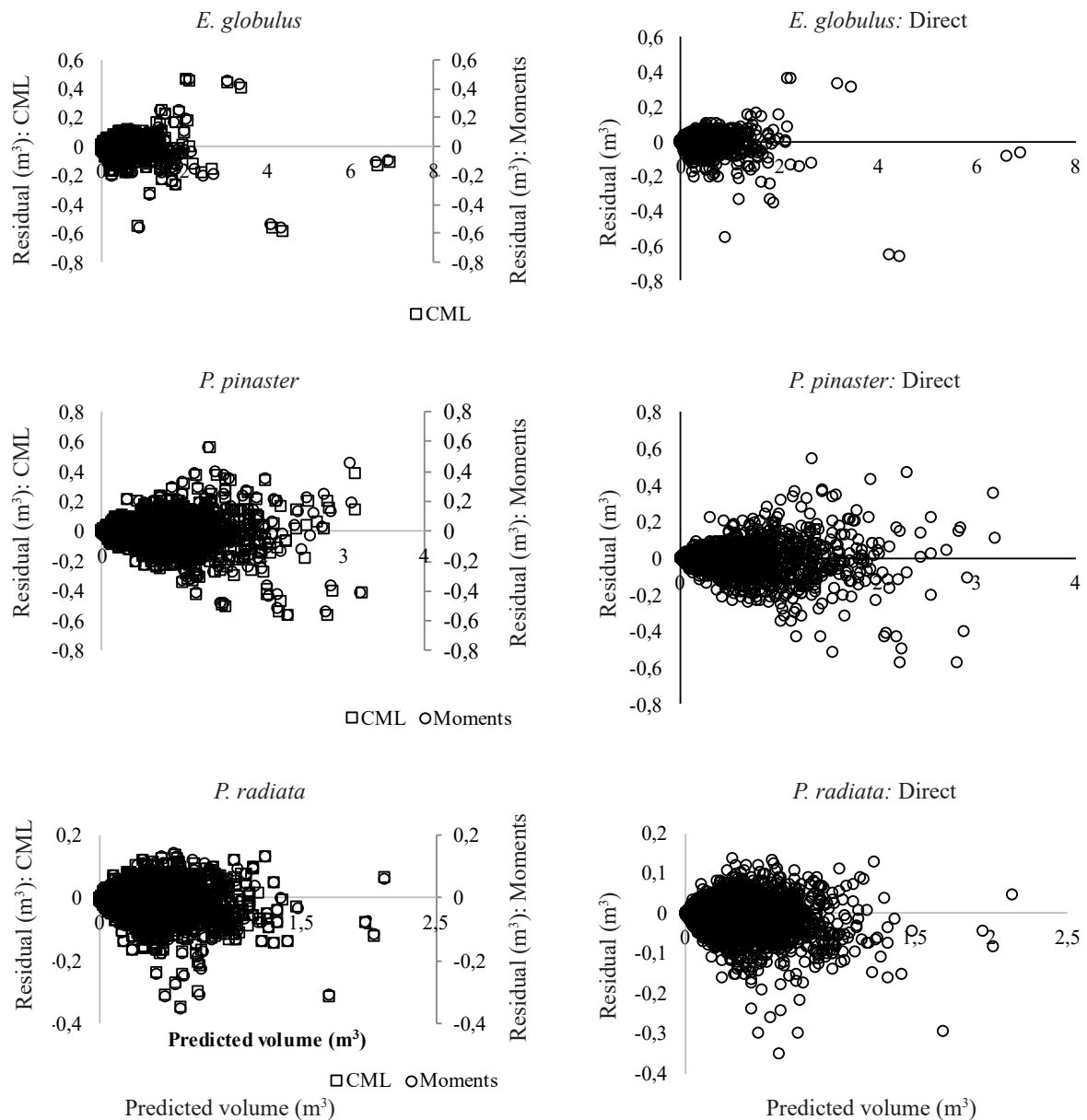
Figure 2. Residual against predicted volume of CML, moments and direct methods across the three species.

estimation. Its accuracy is affected by the method of estimating the parameters of the distribution. In this study, we found the indirect (two-stage) method, especially by conditional maximum likelihood, to be the most suitable method for the bivariate Johnson's S_{BB} distribution. This method predicted tree height and volume in *E. globulus, P. pinaster* and *P. radiata* better than the direct method.

References

Burkhart HE, Tomé M, 2012. Modeling Forest Trees and Stands (2nd Ed). Springer Dordrecht Heidelberg New York, 271 pp. https://doi.org/10.1007/978-90-481-3170-9

Castedo-Dorado F, Ruiz-Gonzalez AD, Álvarez-González JG, 2001. Modelización de la relación altura-diámetro para Pinus pinaster Ait. en Galicia mediante la función de densidad bivariante SBB. Invest Agrar: Sist Recur For 10(1): 111-125.

Clutter JL, Allison BJ, 1974. A growth and yield model for Pinus radiata in New Zealand, pp. 136-160. In: Growth models for tree and stand simulation, Fries J (Ed). Department of Forest Yield Research Note 30, Royal College of Forestry, Stockholm, Sweden.

Diéguez-Aranda U, Rojo-Alboreca A, Castedo-Dorado F, Ávarez-González JG, Barrio-Anta M, Crecente-Campo F, González- González JM, Pérez-Cruzado C, Rodríguez-Soalleiro R, López-Sánchez CA, *et al.* 2009. Herramientas selvicolas para la gestión forestall sostenible en Galicia. Conselleria do Medio Rural, Xunta de Galicia. 268 pp.

Gaffrey D, 1996. Sortenorientiertes Bestandeswachstum-Simulationsmodell auf der Basis intraspezifischen, konkurrenzbedingten Einzelbaumwachstums - insbesondere hinsichtlich des Durchmessers - am Beispiel der Douglasie. In: Berichte des Forschungszentrums Waldökosysteme: Reihe A, 133: 413 pp.

García-Villabrille JD, 2015. Modelización del crecimiento y la producción de plantaciones de Eucalyptus globulus Labill. en el NO de Espa-a. Doctoral thesis. Universidad de Santiago de Compostela, Spain. [in Spanish].

Gorgoso-Varela JJ, Garcia-Villabrille JD, Rojo-Alboreca A, von Gadow K, Alvarez-Gonzalez JG, 2016. Comparing Johnson's SBB, Weibull and Logit-Logistic bivariate distributions for Modeling tree diameters and heights using copulas. Forest Syst, 25(1) 1-5. https://doi.org/10.5424/fs/2016251-08487

Gorgoso JJ, Rojo A, Camara-Obregon A, Dieguez-Arenda 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 Syst, 21(3): 446-459. https://doi.org/10.5424/fs/2012213-02736

Hafley WL, Buford MA, 1985. A bivariate model for growth and yield prediction. For Sci, 31: 237-247.

Johnson NL, 1949a. Systems of frequency curves generated by methods of translation. Biometrika, 36: 149-176. https://doi.org/10.1093/biomet/36.1-2.149

Johnson NL, 1949b. Bivariate distributions based on simple translation systems. Biometrika, 36:297-304. https://doi.org/10.1093/biomet/36.3-4.297

Kalbi S, Fallah A, Bettinger P, Shataee S, Yousefpour R, 2017. Mixed-effects modelling for tree height prediction models of Oriental beech in the Hyrcanian forets. J For Res. https://doi.org/10.1007/s11676-017-0551-z

Knoebel BR, Burkhart HE, 1991. A bivariate distribution approach to modelling forest diameter distributions at two points in time. Biometrics, 47: 241-253. https://doi.org/10.2307/2532509

Li F, Zhang L, Davis CJ, 2002. Modelling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. For Sci, 48(1): 47-58.

Mayer DG, Butler DG, 1993. Statistical validation. Ecol Model, 68:21-32. https://doi.org/10.1016/0304-3800(93)90105-2

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, 2015. The bivariate power-normal and the bivariate Johnson's system bounded distribution in forestry, including height curves. Can J For Res, 45: 307-313. https://doi.org/10.1139/cjfr-2014-0333

Nelder JA, Mead R, 1965. A simplex algorithm for function minimization. Computer Journal, 7: 308-313. https://doi.org/10.1093/comjnl/7.4.308

Ogana FN, 2018a. Evaluation of four methods of fitting Johnson's SBB for height and volume predictions. J For Sci, 64: 187 - 197. https://doi.org/10.17221/151/2017-JFS

Ogana FN, 2018b. Comparison of a modified log-logistic distribution with established models for tree height prediction. J Res For Wild Environ, 10(2): 49 - 55.

Ogana FN, Osho JSA, Gorgoso-Varela JJ, 2018. An approach to modelling the joint distribution of tree diameter and height data. J Sustain For, 37(5):475-488. https://doi.org/10.1080/10549811.2017.1422434

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Schreuder HT, Hafley WL, 1977. A useful bivariate distribution for describing stand structure of tree heights and diameters. Biometrics, 33:471-478. https://doi.org/10.2307/2529361

Siipilehto J, 1996. Metsikön läpimitta- ja pituusjakauman kuvaaminen kaksiulotteisen todennäköisyysfunktion avulla. Licentiate's thesis. University of Helsinki.

Siipilehto J, 2000. A comparison of two parameter prediction methods for stand structure in Finland. Silva Fenn 34(4): 331-349. https://doi.org/10.14214/sf.617

Tewari VP, Singh B, 2018. Total wood volume equation for Tectona grandis Linn F. stands in Gujarat, India. J For Environ Sci, 34(4): 313-320.

Tewari VP, von Gadow K, 1999. Modelling the relationship between tree diameters and heights using SBB distribution. For Ecol Manage, 119: 171-176.

Tewari VP, von Gadow K, 1997. Fitting a bivariate distribution to diameter-height data of forest trees. Indian Forester, 123(9): 815-820.

Wang M, Rennolls K, Tang S, 2008. Bivariate distribution modelling with tree diameter and height: dependency modelling using copulas. For Sci 54(3): 284-293.

Wang M, Rennolls K, 2007. Bivariate distribution modeling with tree diameter and height data. For Sci, 53(1):16-24

Wang M, 2005. Distributional Modelling in forestry and remote sensing. PhD thesis, Univ. of Greenwich, UK, 187 pp.

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 Fenn, 35(2): 169-183. https://doi.org/10.14214/sf.594