Scandinavian Journal of Forest Research

ISSN: 0282-7581 (Print) 1651-1891 (Online) Journal homepage: www.tandfonline.com/journals/sfor20

Modeling diameter distributions of mixed-species forest stands

Fuxiang Liu, Fengri Li, Lianjun Zhang & Xingji Jin

Scandinavian Journal of Forest Research, 2014
Vol. 29, No. 7, 653–663, http://dx.doi.org/10.1080/02827581.2014.960891

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Modeling diameter distributions of mixed-species forest stands

Fuxiang Liu[a,b], Fengri Li[a*], Lianjun Zhang[c] and Xingji Jin[a]

[a]*Department of Forest Management, School of Forestry, Northeast Forestry University, Harbin, Heilongjiang 150040, PR China;* [b]*Department of Statistics, School of Economics, Harbin University of Commerce, Harbin, Heilongjiang 150028, PR China;* [c]*Department of Forest and Natural Resources Management, College of Environmental Science and Forestry (SUNY-ESF), State University of New York, One Forestry Drive, Syracuse, NY 13210, USA*

(*Received 1 July 2013; accepted 29 August 2014*)

This study compares three diameter distribution models to fit mixed-species forest stands using four example plots with two or three species components in Daxing'an Mountain, PR China. The methods include (1) a finite mixture model (FMM) to fit two or three species components simultaneously, (2) a single Weibull function to fit the whole plot only, and (3) a single Weibull function to fit each species component separately and the summation of the individual species produced the whole plot. Our results indicated that Method 2 is only suitable to regular and unimodal diameter distributions with a balanced reversed J-shape. Method 3 may be able to fit each species component well if its frequency distribution is known and available in the data. However, Method 3 ignores the interspecies relationships within a given plot. Thus, the summation of the species components may not produce a good fit for the whole plot. In contrast, Method 1 (FMM) fits the species component distributions simultaneously with the constraint that the individual components add up to the whole plot, without requiring the observed frequencies for each species across the diameter classes. The FMM models are more flexible to describe highly skewed and irregular diameter distributions for the whole plot, as well as provide the acceptable estimation for each species component and the mixing proportions. Thus, the FMM models can be a useful tool for effectively managing mixed-species forest stands.

Keywords: diameter distribution; mixed-species forest stand; Weibull function; finite mixture model

Introduction

Diameter distribution models characterize the age and size structure and tree-layer profile of forest stands. They have been a useful tool for predicting stand growth and yield, updating forest inventory, and planning forest management activities such as thinning or rotation. Over the last decades, a variety of probability density functions (*pdf*) have been used to model diameter distributions, such as normal, beta, gamma, lognormal, logit-logistic, Johnson's S_B, and Weibull (e.g. Bailey & Dell 1973; Little 1983; Maltamo et al. 1995; Wang & Rennolls 2005; Stankova & Zlatanov 2010). Each of these statistical distribution functions has strengths and weaknesses, and may or may not be able to adequately fit a given data-set, depending on many factors such as stand age, stand structure (even-aged, uneven-aged, or irregular), species composition (single or multiple species), etc. Studies have attempted to compare and identify which statistical function is more suitable to describe a given diameter distribution and/or which parameter estimation method is superior to fit a particular statistical function. Hafley and Schreuder (1977) compared six statistical functions including beta, normal, lognormal, gamma, Johnson's S_B, and Weibull in terms of their flexibility and concluded that Johnson's S_B was the best for fitting a variety of stand size distributions. However, Johnson's S_B is more complex and relatively difficult to apply in practice. Therefore, researchers and practitioners prefer a statistical function for which parameter estimation is simple enough to fit the size distributions with different shapes and degrees of skewness, and for which the cumulative density function is easily obtained. Weibull function has been well known as a good choice for fitting unimodal distributions of single species and even-aged stands due to its simplicity, flexibility, and applicability (e.g. Newton et al. 2005; Castedo-Dorado et al. 2007).

Different parameter estimation methods have been applied to fit these statistical distribution functions, such as maximum likelihood (ML), moments, percentiles, and nonparametric (Bayesian and kernel) methods (e.g. Borders & Patterson 1990; Maltamo et al. 2000). Zhang et al. (2003) compared different estimation methods for fitting Johnson's S_B and three-parameter Weibull functions to the diameter distributions of red spruce (*Picea rubens* Sarg.) and balsam fir (*Abies balsamea* (L.) Mill) stands, and found that the relative performance of these functions clearly depended on the estimation methods. Some studies calibrated the

*Corresponding author. Email: fengrili@live.com; fengrili@126.com

predictions from diameter distribution models to obtain compatible estimates for other stand characteristics (Gove & Patil 1998; Kangas & Maltomo 2000).

Most diameter distribution models are utilized to fit the distribution of the "whole stand" of a single species, and may not be suitable for mixed-species stands with highly irregular shapes of size distributions. The use of unimodal statistical functions may lead to oversimplified descriptions of these complex stand structures (Maltamo et al. 2000). Over the last decade, forest modelers have attempted to use modern statistical methods and techniques to describe the diameter distributions of multi-age, multi-layer, and multi-species forest stands. Maltamo and Kangas (1998) and Maltamo et al. (2000) utilized percentile prediction and nonparametric statistical models such as kernel estimation and *k*-nearest-neighbor regression to describe multimodal distributions. However, these studies treated each tree species in the mixed stand independently, assuming all species come from the same or similar distribution, and ignored the relationships and differences between species. To simultaneously model the different components (e.g. species) of a complex size distribution, Zhang et al. (2001) applied the finite mixture model (FMM) to demonstrate that it is flexible enough to fit multimodal and highly skewed mixed-species stands, as well as irregular and rotated-sigmoid diameter distributions of uneven-aged, old-growth stands (Liu et al. 2002; Zhang & Liu 2006). Podlaski (2010a) investigated the suitability of two-component Weibull and gamma mixtures for modeling the diameter distribution and compared the methods of choosing initial parameter values. Podlaski (2010b) applied the FMM to describe the distinctly rotated sigmoid type, the bimodal M-shaped type, and the unimodal highly skewed type of diameter breast height distributions for near-natural multilayered *Abies–Fagus* stands. Jaworski and Podlaski (2012) applied FMMs to a mixture of two- and three-component Weibull and gamma models for describing irregular and multimodal diameter distributions.

A frequency distribution composed of two or more component distributions is defined as a "mixture" or "compound" distribution, involving a finite number of components. FMM has been used extensively to fit such distributions in different study fields including medicine, biology, fisheries, forestry, environmental science, engineering, and economics (e.g. Macdonald & Pitcher 1979; Zasada & Cieszewski 2005; Zhang & Liu 2006). FMM simultaneously estimates the parameters of each component distribution in the mixture, as well as the mixing probabilities of component membership (Chandra 1977; Everitt & Hand 1981; Titterington et al. 1985). A variety of statistical distribution functions have been used as the component distribution (Hasselblad 1969; Shaked 1980), such as exponential (Bartholomew 1969), beta (Bremner 1978), logistic (Shah 1963), gamma (Ashton 1971), normal (Behboodian 1970), Weibull (Falls 1970), etc. Statistical methods have been applied to estimate the parameters of mixture distributions, ranging from moment method (Blischke 1962), ML (Falls 1970), graphical technique (Kao 1959) to Bayes estimation (Padgett & Tsokos 1978).

Forest researchers and practitioners used to consider that FMM was too complicated to use and apply. With recent and rapid improvement and advances on computing technology and software packages such as R package (Leisch 2004), Stata (Deb 2008), SAS (SAS Institute Inc. 2010), and others (Haughton 1997), it is now relatively easy to fit diameter distributions by FMM. However, the studies on FMM in the forestry and ecology literature were most focused on comparing the model fitting of "whole stand" by different methods (e.g. Zhang et al. 2001; Zhang and Liu 2006). Little attention has been paid to the fitting of component distributions in the mixture distribution. Therefore, the main purpose of this study was to fit mixed two- and three-species forest stands, using four example plots, by three methods: (1) a FMM model for fitting the two or three species components simultaneously, (2) a single Weibull function for fitting the whole stand, and (3) a single Weibull function for fitting each species component separately. The performance of the three methods was compared with the fitting of the whole stand, the fitting of each species component, and the estimation of the mixing proportions for the species components.

Materials and methods

Theoretical background

Suppose a mixture distribution consisting of k components (j = 1, 2, …, k) with a random variable x of interest under study (e.g. tree diameter). The distribution of the jth individual component is described by a specific *pdf*, $f_j(x)$, and the general *pdf*, $f(x, p)$, for the mixture distribution can be expressed as follows:

$$f(x,p) = \sum_{j=1}^{k} p_j f_j(x) = p_1 f_1(x) + p_2 f_2(x) + \cdots + p_j f_j(x) \cdots + p_k f_k(x) \quad (1)$$

where $p = p_1, p_2, \ldots, p_{k-1}$ is the vector of $k-1$ mixing proportions of the individual components in the mixture and must satisfy the constraints: $0 < p_j < 1$ and $p_k = 1 - \sum_{j=1}^{k-1} p_j$. Note the overall mixture distribution $f(x, p)$ may be a mixture of different component distributions or a mixture of the same component distribution. In this study, the three-parameter Weibull distribution is selected as the component distribution $f_j(x)$:

$$f_j(x) = \left(\frac{c_j}{b_j}\right)\left(\frac{x-a_j}{b_j}\right)^{c_j-1} \exp\left[-\left(\frac{x-a_j}{b_j}\right)^{c_j}\right], \quad x \geq a_j, a_j \geq 0, b_j > 0, c_j > 0 \quad (2)$$

where a_j, b_j, and c_j are the location, scale, and shape parameters of the jth individual component distribution, respectively. The cumulative distribution function (*cdf*) of the jth individual component distribution is:

$$F_j(x) = 1 - \exp\left[-\left(\frac{x - a_j}{b_j}\right)^{c_j}\right] \quad (3)$$

Then, the compound *cdf* is defined as:

$$F(x,p) = \sum_{j=1}^{k} p_j F_j(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots + p_j F_j(x) \cdots + p_k F_k(x) \quad (4)$$

The ML approach is commonly applied to fit a FMM model, due to its attractive statistical properties and relatively easy to use in practice (Chandra 1977). The joint likelihood density function is as follows:

$$L = \prod_{j=1}^{k} f(x,p) = \prod_{j=1}^{k} [p_1 f_1(x) + p_2 f_2(x) + \cdots + p_j f_j(x) \cdots + p_k f_k(x)] \quad (5)$$

The natural logarithm of the likelihood function (logL) is expressed by:

$$\log L = \sum_{j=1}^{k} \log[f(x,\ p)] = \sum_{j=1}^{k} \log[p_1 f_1(x) + p_2 f_2(x) + \cdots + p_j f_j(x) \cdots + p_k f_k(x)] \quad (6)$$

The partial derivatives of log$L$ are taken with respect to each of the model and mixing parameters of the mixture distribution. These partial derivatives are set equal to zero, and then solved by a numerical iterative algorithm such as the Newton–Raphson algorithm to yield the ML estimates (McLachlan & Peel 2000).

Example plots and modeling methods

The four example plots were obtained from the mixed-species natural forest stands located in Mohe County (52.10–53.33 N, 121.07–124.20 E), Daxing′an Mountain, Heilongjiang Province, PR China. Mohe County is one of the highest latitude areas in China. The soil types belong to forest soil, meadow soil, and bog soil. The vegetation types are boreal coniferous forest and coniferous mixed forest. The forest region belongs to cool temperate zone continental monsoon climate. The annual average frost-free period is around 80–110 days. The average temperature is around –2.6°C and the extreme minimum temperature can reach –52.3°C. The major tree species include Dahurian larch (*Larix gmelinii* Rupr.), white birch (*Betula platyphylla* Suk.), and Mongolian pine (*Pinus sylvestris* L. *var. mongolica* Litv.).

The four fixed sample plots were set up in 2000 and remeasured in 2005 and 2010. This study used the data of 2005. The plots were 0.06 ha in size. Trees with diameter at breast height ≥5 cm were measured on the sample plots with an accuracy of 1 mm. The descriptive statistics of tree diameters for the four example plots and each species in the plot were shown in Table 1. All trees were grouped into 2-cm diameter classes. The ages of the four plots were between 60 and 70 years old with basically reversed J-shaped diameter distributions. More trees were gathered in small diameter classes (5–10 cm). The features of the four example plots were (1) Plot 1 had 63% Dahurian larch (5–25 cm) and 37% white birch (5–15 cm). The two species had similar frequency distributions (reversed J-shaped), as well as similar to the distribution of the whole plot; (2) Plot 2 had 67% Dahurian larch (5–17 cm) and 33% white birch (5–37 cm). The frequency distribution of Dahurian larch was highly skewed to the right, while the frequency distribution of white birch was uniform across the diameter classes; (3) Plot 3 had three tree species. Dahurian larch (41%, 5–23 cm) had a relatively normal frequency distribution, white birch (52%, 5–19 cm) had a skewed distribution to the right due to a few large-sized trees, and Mongolian pine (7%, 5–19 cm) had a uniform distribution across the diameter classes; and (4) Plot 4 also had three species: 42% Dahurian larch (5–27 cm), 25% white birch (5–16 cm), and 34% Mongolian pine (5–31 cm). Dahurian larch had a skewed to the right distribution, while both white birch and Mongolian pine had relatively uniform distributions but different ranges of tree sizes.

In this study, we chose the three-parameter Weibull function (Equation [2]) as the component density function. The location parameter a was fixed at 5 cm (i.e. the minimum measured tree diameter) and the scale parameter b and shape parameter **c** were estimated from the data. Each example plot was fitted by three methods:

- Method 1: a FMM (Equation [1]) with the three-parameter Weibull function as the component density function.
- Method 2: a single Weibull function (Equation [2]) to the whole plot (ignoring individual species).
- Method 3: a single Weibull function (Equation [2]) to each individual species separately, and the summation of the individual species produced the whole plot.

The parameters of all methods were obtained by PROC FMM of SAS 9.3 (SAS Institute Inc. 2010). The ML estimation in the procedure was used with the dual quasi-Newton optimization algorithm. Several criteria were computed for model comparison including Akaike information criterion (AIC), bias, square root of mean

Table 1. Descriptive statistics of tree diameters in the four example plots.

Plot	Species	Number of trees	Observed proportion	Mean	Std	Min	Max	Skewness	Kurtosis
1	Total	61		9.02	3.9109	5	23	1.8097	2.9131
	Larch	38	0.63	9.95	4.507	5	23	1.3435	0.9447
	White birch	22	0.37	7.55	1.8022	5	13	1.2076	0.7971
2	Total	111		10.18	6.5074	5	37	2.0726	3.8867
	Larch	74	0.67	7.41	1.9096	5	15	1.6474	2.5911
	White birch	37	0.33	15.73	8.576	5	37	0.6317	−0.5389
3	Total	100		9.84	3.4604	5	23	0.8763	0.3629
	Larch	41	0.41	8.15	2.943	5	19	1.7107	2.3512
	White birch	52	0.52	11.12	3.2913	5	23	0.7265	0.7588
	Mongolian pine	7	0.07	7.56	1.45	5	19	0.13	0.37
4	Total	89		12.74	6.6999	5	31	1.0555	0.0941
	Larch	37	0.42	11.62	5.479	5	27	1.1369	0.3965
	White birch	22	0.25	9.64	2.9931	5	15	0.1493	−1.4758
	Mongolian pine	30	0.34	16.4	8.2203	5	31	0.2284	−1.3552

squared error (RMSE), and likelihood ratio χ^2 test as follows:

$$\mathrm{AIC} = -2\log L + 2m$$

where logL is the log-likelihood of the model, and m is the number of effective parameters. The AIC is designed as the smaller, the better:

$$\mathrm{Bias} = \frac{\sum_{i=1}^{S}\left(D_i - \hat{D}_i\right)}{S}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{S}\left(D_i - \hat{D}_i\right)^2}{S}}$$

where S is the number of diameter classes, D_i is the diameter sum of the ith diameter class, and $\hat{D}_i$ is the estimated diameter sum by the fitted models. The diameter sum for the ith diameter class is calculated by $D_i = \sum_{i=1}^{S} N_i d_i$(Maltamo et al. 1995), where N_i is the number of trees measured in the ith diameter class, and d_i is the middle value of the ith diameter class:

$$\text{Likelihood ratio } \chi^2 = -2\sum_{i=1}^{S} N_i \log\left(\frac{\hat{N}_i}{N_i}\right)$$

where N_i is the observed frequency (number of trees) of the ith diameter class, and $\hat{N}_i$ is the predicted frequency (number of trees) of the ith diameter class from the fitted models. The χ^2 has (N–m–1) degrees of freedom, where m is the number of estimated parameters.

In this study, the model residual was defined as the difference between the observed diameter sums of the ith diameter class D_i and the predicted diameter sums of the ith diameter class $\hat{D}_i$ by the fitted model. Thus, positive residuals represent underestimation and negative residuals represent overestimation by the model. As Maltamo et al. (1995) indicated that this model residual $e_i = D_i - \hat{D}_i$ puts heavier weights on valuable larger-sized trees.

Results

Model fitting of whole plot

The parameter estimates of the three fitting methods and the standard errors (in parentheses) of the parameters for the four example plots were presented in Table 2. It was evident, according to the model AIC, that Method 1 (FMM) fitted each example plot better than the Method 2 (a single Weibull) to fit the whole plot, except for Plot 3 (Table 2). However, the Method 1 (−2log$L$ = 493.4) fitted the Plot 3 better than the Method 2 according to the likelihood value (−2log$L$ = 498.3), but the AIC is designed to penalize the number of estimated model parameters. Because the number of the FMM model parameters (m = 8) was much larger than those of the single Weibull model parameters (m = 2), the AIC of the Method 1 (509.4) was larger than that of the Method 2 (502.3). Note there was no AIC value for the Method 3 because it fitted each species separately and no model fitting statistics were available for the whole plot.

To evaluate the model fitting and prediction performance, the bias, RMSE, and goodness-of-fit likelihood ratio χ^2 test were computed and shown for each method and each example plot in Table 3. The model prediction and model residuals across the diameter classes are displayed for each example plot in Figures 1 and 2, respectively. The Plot 1 had a balanced reversed J-shape diameter distribution for the whole plot (Figure 1a). Based on the likelihood ratio χ^2 test all three methods

Table 2. Parameter estimates (standard error) of the three model fitting methods for the four example plots.

Plot	b_1	c_1	b_2	c_2	b_3	c_3	p_1	p_2	AIC
Method 1									
1	2.395 (0.3069)	0.5954 (0.1157)	8.3111 (0.5536)	0.6048 (0.3283)			0.6424 (0.4413)		292.8
2	2.0881 (0.1284)	0.5775 (0.06948)	12.4659 (0.2406)	0.6755 (0.1781)			0.6456 (1.4309)		562.6
3	7.2498 (2.0595)	0.5439 (0.4586)	2.3807 (0.8340)	0.5468 (0.09926)	0.8674 (0.03490)	1.828822 (0.09926)	0.3396 (12.7683)	0.4136 (6.2704)	509.4
4	4.3579 (0.3276)	0.1141 (0.02809)	11.1591 (0.1555)	0.1723 (0.1326)	21.8590 (0.3276)	0.7031 (0.09881)	0.5633 (1.2657)	0.1761 (0.4325)	539.9
Method 2									
Plot	b	c							AIC
1	4.2359 (0.1191)	0.8760 (0.08295)							293.8
2	4.9006 (0.1101)	1.0912 (0.07586)							589.3
3	5.3374 (0.07411)	0.7021 (0.05538)							502.3
4	8.1414 (0.09755)	0.8713 (0.07256)							543.7
Method 3									
Plot	b_1	c_1	b_2	c_2	b_3	c_3			
1	5.223 (0.1491)	0.8676 (0.1074)	2.8486 (0.1483)	1.433551 (0.1052)					
2	2.6673 (0.08814)	0.7134 (0.06071)	11.1856 (0.1531)	0.886 (0.1201)					
3	6.8959 (0.07538)	0.5163 (0.05561)	3.3708 (0.1385)	0.8346 (0.09575)	8.0690 (0.9415)	0.3046 (0.7027)			
4	7.1209 (0.1401)	0.8059 (0.1026)	5.1399 (0.1474)	0.6572 (0.1166)	12.1095 (0.1565)	0.8177 (0.1261)			

adequately fitted this plot (p-value > 0.05), although Method 1 had much smaller bias and RMSE than those of other two methods (Table 3). Methods 2 and 3 underestimated for the diameter classes 8 and 22 cm and overestimated for other diameter classes (Figure 2a). In contrast, the Plot 2 had highly skewed diameter distribution (Figure 1b). Only Method 1 fitted the data well, while both Methods 2 and 3 did not adequately fit this plot (p-value < 0.05). Both methods underestimated for the diameter class 8 cm and the classes >20 cm, but overestimated for the diameter classes 10, 12, and 16 (Figure 2b). Method 3 fitted the plot better than Method 2 because a single Weibull function (Method 2) may not be flexible enough to fit a highly skewed frequency distribution.

Both Plots 3 and 4 had three species components and the diameter distributions of the whole plot were more irregular in shape, either bimodal (Plot 3, Figure 1c) or rotated-sigmoid (Plot 4, Figure 1d). For both plots, Method 1 (FMM) was the only one fitting these irregular distributions well, while both Methods 2 and 3 failed to adequately describe them according to the likelihood ratio χ^2 tests (p-value < 0.05) and yielded much larger biases and RMSEs (Table 3).

Model fitting of individual species components

The comparison between Methods 1 and 3 for fitting the frequency distribution of each individual species is shown in Figure 3 and Table 4. The two species (larch and white birch) of the Plot 1 had similar frequency distributions – balanced reversed J-shape (Figure 3a and 3b, respectively). Method 3 fitted these two distributions better (smaller bias and RMSE) than Method 1 (Table 4), especially for small diameter classes. In contrast, the two species of Plot 2 had different frequency distributions in shape, e.g. the distribution of larch was highly skewed, while the distribution of white birch was uniform across the diameter classes. In this case, Method 1 fitted better than Method 3 for each component with smaller bias and RMSE (Figure 3c and 3d, respectively, and Table 4), indicating that FMM is more suitable for fitting heterogeneous diameter distributions.

Plot 3 had three tree species. The distribution of larch was relatively normal (Figure 3e), the distribution of white birch was skewed to the right due to a few large-sized trees (Figure 3f), and the distribution of Mongolian pine was uniform across the diameter classes (Figure 3g). Method 1 fitted the first component distribution (larch) much better than Method 3; while Method 3 fitted the other species better, especially Mongolian pine which had only seven trees in five diameter classes. It seemed that the FMM model had a hard time to fit this species component. In Plot 4 larch had a skewed to the right distribution (Figure 3h), while both white birch (Figure 3i) and Mongolian pine (Figure 3j) had relatively uniform distributions but different ranges of

Table 3. Bias, RMSE, and χ^2 test of the three model fitting methods for the four example plots.

	Method 1				Method 2				Method 3			
Plot	Bias	RMSE	χ^2	p-value	Bias	RMSE	χ^2	p-value	Bias	RMSE	χ^2	p-value
1	−0.0001	0.00026	0.0001	1.00	1.4669	19.1883	8.3725	0.2121	4.1682	13.3194	8.8278	0.0656
2	−0.0001	0.00014	0.0001	1.00	3.3003	47.4122	40.174	0.0007	3.9480	28.6135	19.845	0.0306
3	0.0001	0.00022	0.0002	0.98	23.2515	63.8711	49.975	0.0001	30.1097	76.2834	62.712	0.0001
4	0.0001	0.00060	0.0002	1.00	17.6735	31.2182	25.600	0.0023	0.6140	37.0708	34.300	0.0001

tree sizes. Method 1 fitted the distribution of Mongolian pine better, but Method 3 fitted the other two species better (Table 4). However, the FMM model compromised the three species to obtain a good fit for the whole plot (Figure 1d).

Estimation of component proportions

The FMM model (Method 1) estimates the component proportions simultaneously while fitting the component distributions. Again, the only information needed is the number of species components in each plot. On the other hand, Method 3 is fitted to the known component distributions first, and the components are added up to the whole plot. Then, the component proportion for each species can be computed. Table 4 showed the observed component proportions and estimated proportions by Methods 1 and 3. The number in the parentheses in the columns of "Estimated Proportions" represented the difference between the observed proportions and estimated proportions. The positive numbers represented underestimation, while negative numbers represented overestimation.

For Plots 1 and 2, both Methods 1 and 3 estimated the two component proportions equally well, which were very close to the observed proportions. For Plot 3, Method 1 slightly underestimated larch, while Method 3

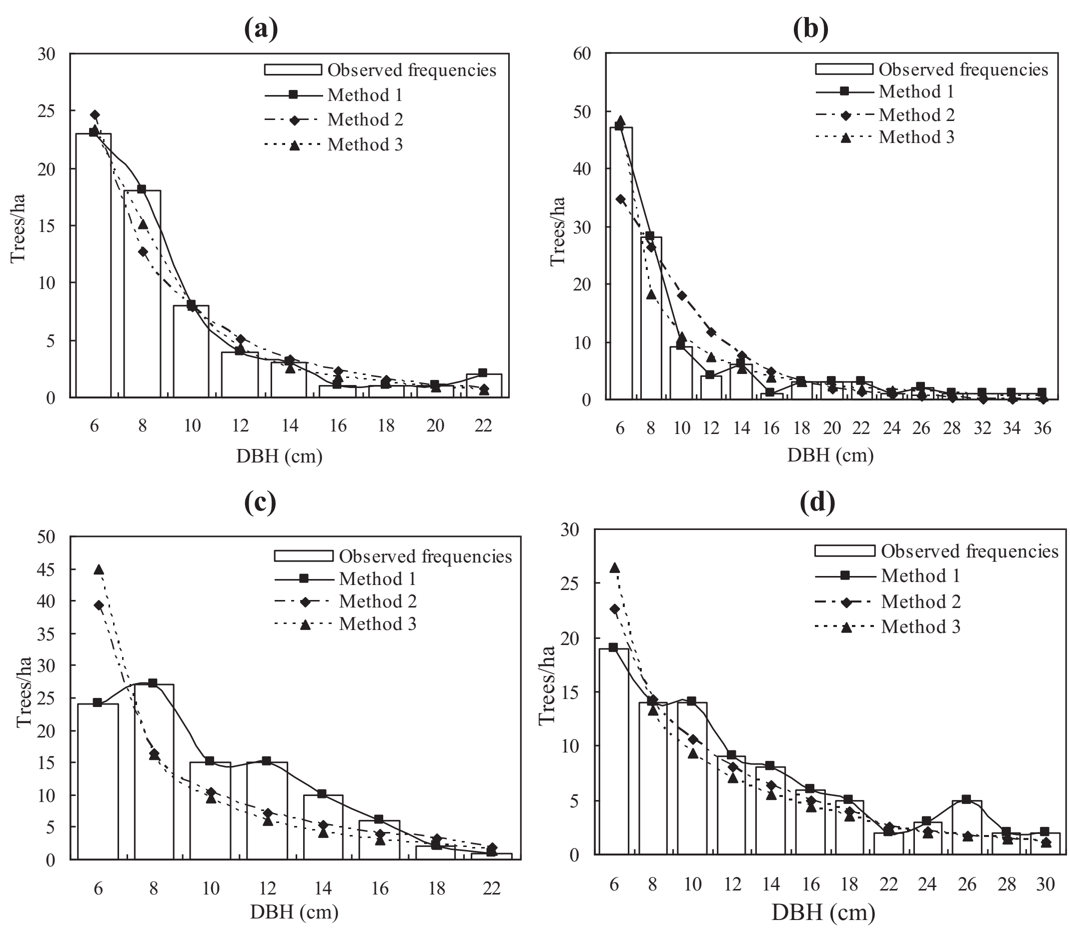

Figure 1. Comparison of three model fitting methods for the example plots. (a) Plot 1, (b) Plot 2, (c) Plot 3, and (d) Plot 4.

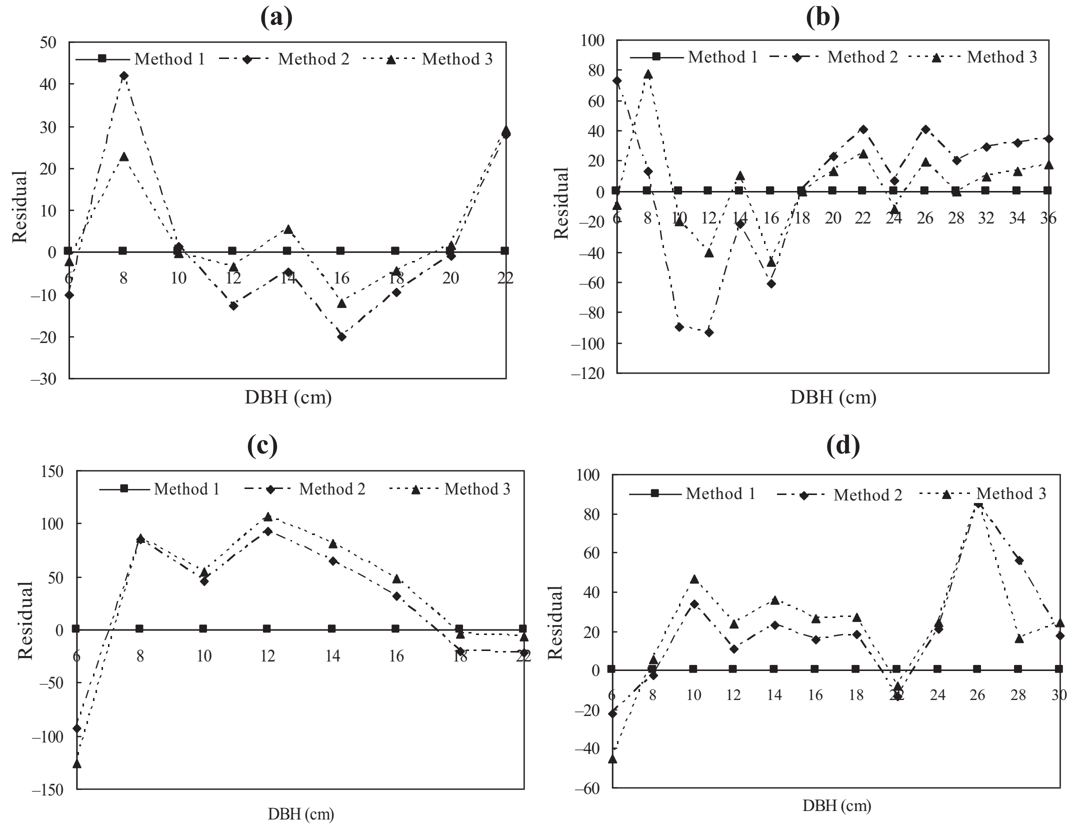

Figure 2. Comparison of model residuals (sum diameters) for the example plots. (a) Plot 1, (b) Plot 2, (c) Plot 3, and (d) Plot 4.

largely overestimated this species (due to the overestimation for small-sized trees (Figure 3e). Both methods underestimated white birch. However, Method 1 greatly overestimated Mongolian pine, but Method 3 did a good job on estimating its proportion (Table 4). Because the FMM model happened to underestimate the proportions (p_1 and p_2) of the other two species, the estimated proportion of Mongolian pine was equal to ($1–p_1–p_2$) resulting in overestimating its proportion. For Plot 4, both Methods 1 and 3 produced compatible estimation on the proportions of the three species (Table 4).

Discussion

Model fitting of whole plot

The objective of this study was to compare three diameter distribution models to fit mixed-species forest stands with two or three species components, in order to predict diameter distributions of the whole plots from typical inventory data, as well as to predict diameter distributions of individual species component. Our model fitting and comparison results indicated that the mixture Weibull models (Method 1) were more flexible to fit regular and irregular diameter distributions, while a single Weibull (Method 2) and the sum of component Weibull functions (Method 3) may not be able to adequately describe highly skewed or irregular distributions of the whole plots (Figure 1, Table 3). Methods 2 and 3 produced typical reversed J-shape curves for these two plots, thus overestimated for small diameter classes and underestimated for larger diameter classes (Figure 2c,d, respectively). Our results confirmed the findings in previous studies (e.g. Zhang et al. 2001; Liu et al. 2002; Zhang & Liu 2006; Jaworski & Podlaski 2012). Therefore, the finite mixture distribution (Method 1) is a better method for modeling the diameter distributions of mixed-species forest stands with two or more species components (Table 3; Figure 1).

Model fitting of individual species components

Both Methods 1 and 3 can estimate the frequency distributions of individual species component in the mixed-species plots, but in very different ways. Method 1 (FMM) (Equation [1]) fits the two or three component distributions simultaneously with the constraint that the individual components add up to the whole plot. However, the observed frequencies for each species across the diameter classes are not required. The only information needed is the number of species components in each plot. In contrast, Method 3 fits a single Weibull function (Equation [2]) to the frequency distribution of each species separately and the mixing connection of these

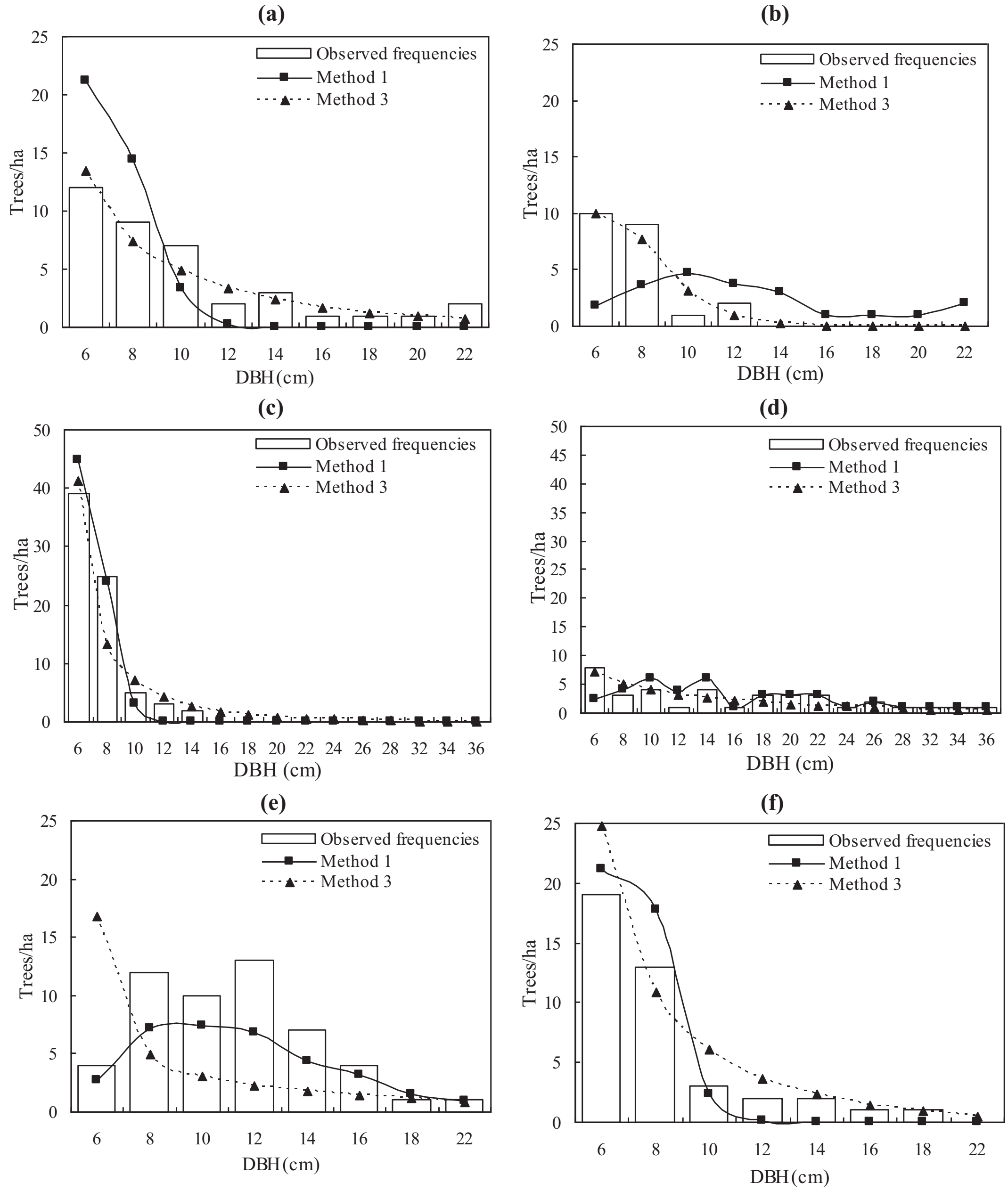

Figure 3. Comparison of model fitting of each species for the example plots. Plot 1: (a) Larch, (b) white birch; Plot 2: (c) larch, (d) white birch; Plot 3: (e) larch, (f) white birch, (g) Mongolian pine; and Plot 4: (h) larch, (i) white birch, (j) Mongolian pine.

species is not considered or constrained. In this case, the observed frequencies for each species across the diameter classes must be known in order to obtain the Weibull parameters for each species. Otherwise, Method 3 cannot be applied to fit each species component of the mixed-species plots.

Method 3 may have an advantage of fitting each species component if its frequency distribution is known and has a relatively regular shape (e.g. balanced reversed J-shape or uniform). However, Method 3 treats each species component separately as a single distribution and never considers the interspecies relationships within a given plot. Thus, the summation of the species components may not produce a good fit for the whole plot. The advantages of the FMM models include not only great flexibility over a single Weibull function, but also avoiding the necessity of classifying two or more components of multimodal distribution a priori during data collection. However, the number of component distributions must be decided when specifying the underlying FMM model to be fitted. This can be done in an exploratory way by fitting several finite mixtures and settling on the one with the best model fitting statistics (Zhang et al. 2001).

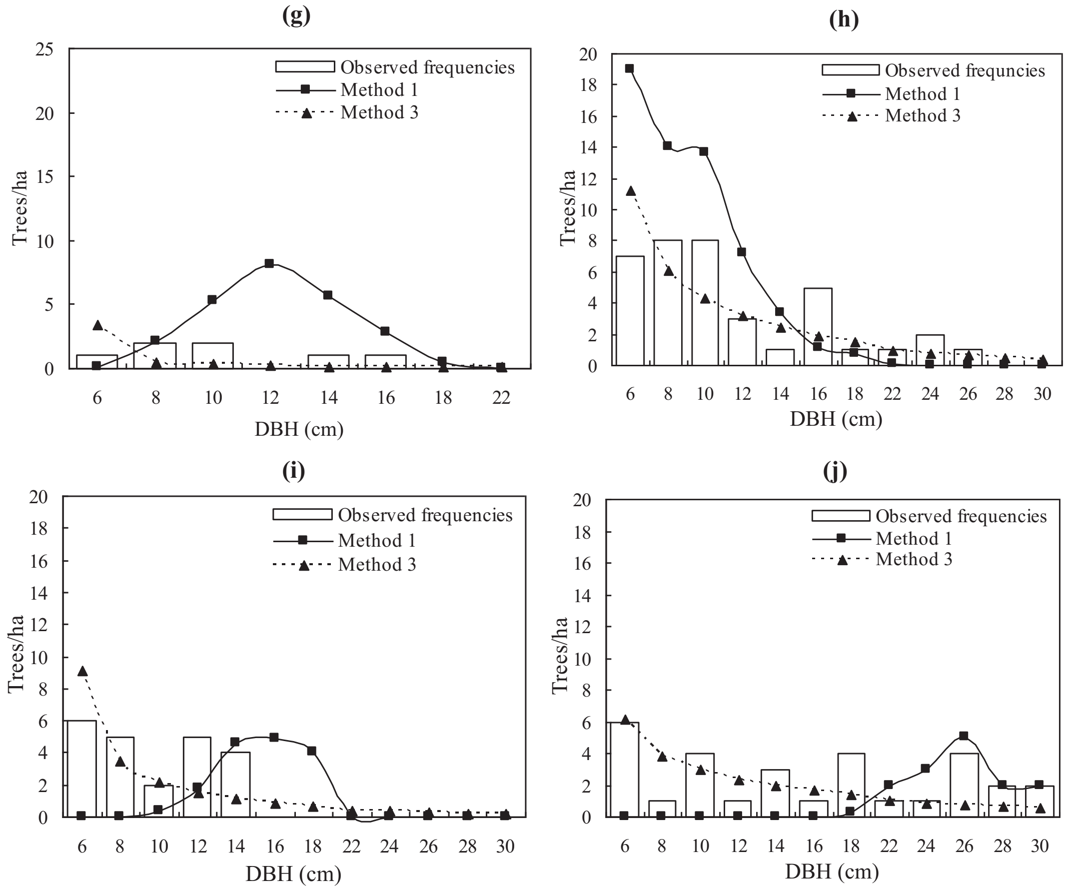

Figure 3. (*Continued*)

Estimation of component proportions

The FMM models provided the acceptable estimation on the proportions of species components in the mixed two or three species plots (Table 4). Note this estimation was based on unknown frequency distributions of different species components. The only information needed was the number of species in the plot, while the actually observed frequency of each species is not required. This feature of the FMM models is particularly important for managing the mixed-species forest stands. In practice, forest managers need the information of diameter distributions on both "whole stand" and individual species

Table 4. Bias, RMSE, and estimated proportions of each species of the Methods 1 and 3 for the four example plots.

			Method 1			Method 3		
Plot	Species	Observed proportion	Bias	RMSE	Estimated[a] proportion	Bias	RMSE	Estimated[a] proportion
1	Larch	0.63	10.9962	1257.83	0.64 (−0.01)	3.7022	225.44	0.62 (+0.01)
	White birch	0.37	−11.6629	1188.64	0.36 (+0.01)	−0.2006	81.284	0.38 (−0.01)
2	Larch	0.67	3.8437	242.69	0.65 (+0.02)	−4.6201	757.79	0.69 (−0.02)
	White birch	0.33	−3.8437	242.69	0.35 (−0.02)	8.5681	398.531	0.31 (+0.02)
3	Larch	0.41	5.0366	1183.00	0.34 (+0.07)	36.9573	4725.47	0.58 (−0.17)
	White birch	0.52	23.7044	438.73	0.41 (+0.11)	−10.9711	376.646	0.37 (+0.15)
	Mongolian pine	0.07	−28.7408	1968.12	0.25 (−0.18)	4.1236	120.33	0.06 (+0.01)
4	Larch	0.42	−8.2546	1795.39	0.56 (−0.14)	5.1480	547.85	0.47 (−0.05)
	White birch	0.25	−2.3338	1332.14	0.18 (+0.07)	1.0681	377.513	0.25 (+0.00)
	Mongolian pine	0.34	10.5885	1076.79	0.26 (+0.08)	15.7108	1142.86	0.28 (+0.06)

[a]The number in the parentheses represents the difference between the observed proportion and estimated proportion.

component in order to effectively conduct silvicultural treatments to achieve the best stand structure and valuable timber for the desirable species. When the detailed inventory data for different species are not readily available the FMM models can provide accurate prediction for the whole plot, as well as acceptable estimations on the species component distributions and the mixing proportions.

Conclusion

Traditional diameter distribution models commonly apply a statistical function (e.g. Weibull) to fit the frequency distribution of the whole plot. These models may be sufficient for fitting unimodal distributions of single species and even-aged stands. For mixed-species stands, a Weibull function can be used to fit the distribution of each individual species separately, when the detailed inventory data are available for these species. This method may be able to fit each species component well, but the summation of the species components may not produce a good fit for the whole plot because it ignores the interspecies relationships within the stand.

The FMM fits the species component distributions in a mixed-species stand simultaneously with the constraint that these individual components add up to the whole plot, without requiring observed frequencies for each species across the diameter classes. The only information needed is the number of species components in each plot. The results in this study confirmed that the FMM models are more flexible to describe highly skewed and irregular diameter distributions for the whole plot, as well as provide the acceptable estimation for each species component and the mixing proportions. In theory, the FMM model is capable of modeling a finite number of individual components in a mixture distribution. But the more individual components need to be modeled, the larger data-set is required in order to have sufficient trees for each species. Otherwise, the FMM model may produce inaccurate estimation for the individual components that have only a few trees available in the stand. In this case, forester managers may use their professional knowledge to combine minor species into species groups. In any case, the FMM models can be a useful tool for effectively managing mixed-species forest stands.

Funding

This research was financially supported by the Ministry of National Science and Technology, PR China (Project ID 2012BAD22B0202) and the Scientific Research Funds for Forestry Public Welfare of China (Project ID 201004026).

References

Ashton WD. 1971. Distribution for gaps in road traffic. IMA J Appl Math. 7:37–46.

Bailey RL, Dell TR. 1973. Quantifying diameter distributions with the Weibull function. For Sci. 19:97–104.

Bartholomew DJ. 1969. Sufficient conditions for a mixture of exponentials to be a probability density function. Ann Math Stat. 40:2183–2188.

Behboodian J. 1970. On a mixture of normal distributions. Biometrika. 57:215–217.

Blischke WR. 1962. Moment estimators for the parameters of a mixture of two binomial distributions. Ann Math Stat. 33:444–454.

Borders BE, Patterson WD. 1990. Projecting stand tables: a comparison of the Weibull diameter distribution method, a percentile-based projection method, and a basal area growth projection method. For Sci. 36:413–424.

Bremner JM. 1978. Algorithm AS 123: mixtures of beta distributions. Appl Stat. 27:104–109.

Castedo-Dorado F, Diéguez-Aranda U, Álvarez-González JG. 2007. A growth model for *Pinus radiata* D. Don. stands in north-western Spain. Ann For Sci. 64:453–465.

Chandra S. 1977. On the mixtures of probability distributions. Scand J Stat. 4:105–112.

Deb P. 2008. Finite mixture models. Summer North American Stata Users' Group Meetings 2008 7, Stata Users Group. [revised 2008 Aug 28; cited 2013 Nov 1]. Available from: www.stata.com/meeting/snasug08/deb_fmm_slides.pdf

Everitt BS, Hand DJ. 1981. Finite mixture distributions. Cambridge (UK): Cambridge University Press; p. 143.

Falls LW. 1970. Estimation of parameters in compound Weibull distributions. Technometrics. 12:399–407.

Gove JH, Patil GP. 1998. Modeling basal area-size distribution of forest stands: a compatible approach. For Sci. 44: 285–297.

Hafley WL, Schreuder HT. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Can J For Res. 7:481–487.

Hasselblad V. 1969. Estimation of finite mixtures of distributions from the exponential family. J Am Stat Assoc. 64: 1459–1471.

Haughton D. 1997. Packages for estimating finite mixtures: a review. Am Stat. 51:194–205.

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

Kangas A, Maltomo M. 2000. Calibrating predicted diameter distribution with additional information. For Sci. 46: 390–396.

Kao JHK. 1959. A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. Technometrics. 1:389–407.

Leisch F. 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. Universitat fur Bodenkultur Wien. [cited 2013 Nov 1]. Available from: cran.r-project.org/web/packages/flexmix/vignettes/flexmix-intro.pdf

Little SN. 1983. Weibull diameter distributions for mixed stands of western conifers. Can J For Res. 13:85–88.

Liu C, Zhang L, Davis CJ, Solomon DS, Gove JH. 2002. A finite mixture model for characterizing the diameter distribution of mixed-species forest stands. For Sci. 48: 653–661.

Macdonald PDM, Pitcher TJ. 1979. Age groups from size frequency data: a versatile and efficient method of analyzing distribution mixtures. J Fish Res Board Can. 36:987–1001.

Maltamo M, Kangas A. 1998. Methods based on k-nearest neighbor regression in estimation of basal area diameter distribution. Can J For Res. 28:1107–1115.

Maltamo M, Kangas A, Uuttera J, Torniainen T, Saramaki J. 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. For Ecol Manag. 133:263–274.

Maltamo M, Puumalainen J, Päivinen R. 1995. Comparison of beta and Weibull functions for modeling basal area diameter distribution in stands of *Pinus sylvestris* and *Picea abies*. Scand J For Res. 10:284–295.

McLachlan G, Peel D. 2000. Finite mixture models. New York: Wiley.

Newton PF, Lei Y, Zhang SY. 2005. Stand-level diameter distribution yield model for black spruce plantations. For Ecol Manag. 209:181–192.

Padgett WJ, Tsokos CP. 1978. On Bayes estimation of reliability for mixtures of life distributions. SIAM J Appl Math. 34:692–703.

Podlaski R. 2010a. Two-component mixture models for diameter distributions in mixed-species, two-age cohort stands. For Sci. 56:379–390.

Podlaski R. 2010b. Diversity of patch structure in Central European forests: are tree diameter distributions in near-natural multilayered Abies-Fagus stands heterogeneous? Ecol Res. 25:599–608.

SAS Institute, Inc. 2010. SAS/STAT user's guide, version 9.3. Cary (NC): SAS Institute.

Shah BK. 1963. A note on the method of moments applied to a mixture of two logistic populations. J MS Univ Baroda (Sci Num). 12:21–22.

Shaked M. 1980. On mixture from exponential families. J R Stat Soc B. 42:192–198.

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.

Titterington DM, Smith AFM, Makov UE. 1985. Statistical analysis of finite mixture distributions. New York: Wiley.

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

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 stand in Poland. For Ecol Manag. 204:145–158.

Zhang L, Gove JH, Liu C, Leak WB. 2001. A finite mixture distribution for modeling the diameter distribution 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. J For Res. 11:369–372.

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.