

Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species: A two-component mixture model approach



Rafał Podlaski^{a,*}, Francis A. Roesch^b

^a Department of Nature Protection, Institute of Biology, Jan Kochanowski University, ul. Świętokrzyska 15, 25-406 Kielce, Poland

^b USDA Forest Service, Southern Research Station, 200 WT Weaver Blvd., Asheville, NC 28804-3454, USA

ARTICLE INFO

Article history:

Received 13 February 2013

Received in revised form 14 January 2014

Accepted 20 January 2014

Available online 31 January 2014

Keywords:

Two-component mixture model

Parameter estimation

Initial values

Weibull distribution

Gamma distribution

Kernel density estimator

ABSTRACT

In recent years finite-mixture models have been employed to approximate and model empirical diameter at breast height (DBH) distributions. We used two-component mixtures of either the Weibull distribution or the gamma distribution for describing the DBH distributions of mixed-species, two-cohort forest stands, to analyse the relationships between the DBH components, age cohorts and dominant species, and to assess the significance of differences between the mixture distributions and the kernel density estimates. The data consisted of plots from the Świętokrzyski National Park (Central Poland) and areas close to and including the North Carolina section of the Great Smoky Mountains National Park (USA; southern Appalachians). The fit of the mixture Weibull model to empirical DBH distributions had a precision similar to that of the mixture gamma model, slightly less accurate estimate was obtained with the kernel density estimator. Generally, in the two-cohort, two-storied, multi-species stands in the southern Appalachians, the two-component DBH structure was associated with age cohort and dominant species. The 1st DBH component of the mixture model was associated with the 1st dominant species *sp1* occurred in young age cohort (e.g., sweetgum, eastern hemlock); and to a lesser degree, the 2nd DBH component was associated with the 2nd dominant species *sp2* occurred in old age cohort (e.g., loblolly pine, red maple). In two-cohort, partly multilayered, stands in the Świętokrzyski National Park, the DBH structure was usually associated with only age cohorts (two dominant species often occurred in both young and old age cohorts). When empirical DBH distributions representing stands of complex structure are approximated using mixture models, the convergence of the estimation process is often significantly dependent on the starting strategies. Depending on the number of DBHs measured, three methods for choosing the initial values are recommended: $\min.k/\max.k$, $0.5/1.5/\text{mean}$, and *multistart*. For large samples (number of DBHs measured ≥ 80) the *multistage* method is proposed – for the two-component mixture Weibull or gamma model select initial values using the $\min.k/\max.k$ (for $k = 1, 5, 10$) and $0.5/1.5/\text{mean}$ methods, run the numerical procedure for each method, and when no two solutions are the same, apply the *multistart* method also.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Management and disturbances play an important role in the forest dynamics as well as in shaping the spatial and dimensional structure of forest stands (e.g., [62,18,37]). After cuttings and disturbances regeneration processes start in the gaps and under open stand canopy; and as a result, among others mixed-species, two-cohort, two-storied and partly multilayered stands are created. These resulting stands are usually characterised by

strongly skewed and irregularly descending diameter at breast height (DBH) distributions. There are two general approaches to fitting empirical DBH distributions. The first approach is nonparametric and therefore does not require the estimation of parameters. The second, and usually preferred, approach is to identify an appropriate parametric distribution, such as e.g., the Weibull or the gamma distribution, and then estimate the unknown parameters. There are several reasons to prefer the latter approach, for instance, nonparametrically binning the data does not provide information beyond the range of the sample data, whereas some extrapolation is possible when a parametric model is applied [7].

* Corresponding author. Tel.: +48 41 3496322; fax: +48 41 3496292.

E-mail address: r_podlaski@pro.onet.pl (R. Podlaski).

When modelling forest dynamics, one must determine the DBH distributions of tree species representing particular cohorts and stand layers (e.g., [5,58,11]), among other parameters. Procedures that allow determination of these parameters from DBH measurements alone, without associated assessments of tree age and height, are particularly valuable. If the overall DBH distribution of a stand is treated as a compound of the distributions of trees belonging to different groups (e.g., cohorts or stand layers) one may adopt a finite-mixture distribution approach (e.g., [73,69,71,46,47]). Mixture distributions are an appropriate tool for modelling heterogeneous populations (e.g., [12,60,35]). Because of their usefulness as an extremely flexible method of modelling, finite mixture models are continuing to receive increasing attention in forestry, from both practical and theoretical points of view. Indeed, in the past decade, the extent and the potential of the applications of finite mixture models have widened considerably [55].

The single and mixture Weibull and gamma models have often been used to approximate empirical DBH distributions because of their flexibility in shape (e.g., [20,32,73,29,72,46,47]). These distributions can conform to a wide variety of DBH data. Because the overall shape of the empirical DBH distribution is often composed of multiple basic shapes, a natural alternative is to utilise a mixture distribution for DBH modelling.

Computation of the parameters for a mixture model can be carried out by various numerical algorithms, such as the expectation–maximisation (EM) algorithm and the Newton-type methods (e.g., [9,36]). The Newton-type methods include quasi-Newton methods, modified Newton methods, etc. These numerical procedures can be decomposed into three main parts: initialisation (in which initial values for all parameters and a criterion to stop the algorithm should be chosen), iteration, and completion when the criterion is met. In general, starting from suitable initial parameter values, the iterations are repeated until convergence is achieved. If the likelihood function is regular, these methods usually find the most likely estimates for mixture parameters. However, if the likelihood function is irregular and has finitely or infinitely many local maxima and minima, the algorithms become extremely unstable. Unfortunately, this concern is a serious obstacle to interpreting the results when applied to separating finite mixtures. Therefore, when the likelihood function is not regular, a combination of the EM algorithm and the Newton-type method is often employed (the EM algorithm improves the initial values, and the Newton-type method is used then to estimate the parameters). In many cases, suitable initial values are difficult to ascertain, especially for empirical DBH distributions representing uneven-aged stands of complex structure. An evaluation of the usefulness of the various methods for choosing the initial values is very important. The appropriate strategies allow one to estimate the parameters of the mixture models and to construct accurate DBH models, especially in difficult situations, such as when the DBH components of mixture models overlap, in which case the global maximum may not be found or the estimation process may fail to converge.

The purposes of this study are (1) to verify the two hypotheses that (a) in mixed-species, two-cohort, two-storied and partly multilayered stands, two-component mixtures of either the Weibull distribution or the gamma distribution would be appropriate models for the DBH distributions; (b) in these models, the DBH components, representing age cohorts (and usually stand layers), can be associated with dominant species; (2) to compare four methods for choosing initial values for the numerical procedure for estimating the parameters of mixture models; (3) to propose a new strategy for maximising the likelihood during parameter estimation for mixture models; and (4) to assess the significance of differences between the parametric (two-component mixture distributions) and the nonparametric (kernel density estimation) methods.

2. Material and methods

2.1. Study area

The plots investigated here were randomly sampled in mixed-species, two-cohort, two-storied and partly multilayered stands in which DBH distributions of two main age cohorts are partially overlapping.

The plots were located in the Święta Katarzyna and Święty Krzyż forest sections of the Świętokrzyski National Park (Poland; Świętokrzyskie Mountains; geographical coordinates: 50°50′–50°53′N, 20°48′–21°05′E); and in areas close to and including the North Carolina section of the Great Smoky Mountains National Park (USA; southern Appalachians; geographical coordinates: 34°59′–36°32′N, 78°43′–84°13′W).

In the Świętokrzyskie Mountains naturally regenerated near-natural forests chosen for this study are composed of native tree species. Soils are Distric Cambisols and Haplic Luvisols (subtypes according to Food and Agriculture Organization, International Soil Reference and Information Centre, and International Soil Science Society, [15]). Long-term mean annual temperature was 6 °C, mean January and July temperatures were –5 °C and 16 °C; the growing season was ca. 182 days (data from the Święty Krzyż meteorological station at 575 m a.s.l.). The highest temperatures and the highest precipitation usually occur in summer, in the middle of the growing season. Three associations occur: *Dentario glandulosae-Fagetum*, *Abietetum polonicum* and *Quercus roboris-Pinetum* (nomenclature after [33]).

In the southern Appalachians, variations in elevation, rainfall, temperature, and geology provide habitat for nearly 1600 species of flowering plants, including 100 native tree species and over 100 native shrub species [38]. Great Smoky Mountains National Park contains some of the largest tracts of wilderness in the Eastern United States, including 66 species of mammals, over 200 varieties of birds, 50 native fish species, and more than 80 types of reptiles and amphibians. The study area is part of the Unaka Range a sub-range in the Appalachian chain, ranging in elevation from about 300–2040 m a.s.l. The climate and precipitation vary greatly in relation to elevation and landscape position. The precipitation averages from 1200 mm annually to approximately 2500 mm at the highest elevations. High precipitation and cool temperatures at the higher elevations produce brown, medium textured soils that have a high content of organic matter in the surface layer. The warmer temperatures at the lower elevations produce soils that are redder and that contain more clay in the subsoil. Mean January temperatures range from 2 °C to 10 °C and mean July temperatures range from 18 °C to 31 °C [61].

2.2. Field measurements

In the Świętokrzyskie Mountains, eleven 0.25 ha plots were measured in 2008 and 2009. In the southern Appalachians, nineteen 0.067 ha plots were measured from 2003 to 2008; data were selected from the USDA Forest Service's Forest Inventory and Analysis database (documented in [68]). The DBH of all live trees greater than 6.9 cm in the Świętokrzyskie Mountains and 12.9 cm in the southern Appalachians in diameter was measured.

2.3. Data analysis

The stands investigated were categorised in three groups. Group 1 (two species stands) consisted of 11 stands that were strongly dominated by two species, from the Świętokrzyski National Park. Group 2 (multi-species stands with two main species) consisted of 10 stands that were medium dominated by two

species, from the southern Appalachians; the basal areas of the two dominant species (main species) were greater than 40% of the total basal area in this group. Group 3 (multi-species stands with no main species) consisted of 9 stands that were weakly dominated by two species, from the southern Appalachians; the basal area of the two most prevalent species did not exceed 40% of the total basal area (Table 1). The proportion of basal area was always largest for the two dominant species (*sp1* and *sp2*) distinguished in each plot investigated. The dominant species *sp1* was most likely associated with the young age cohort, the dominant species *sp2* was initially associated with the old age cohort.

The tree number per plot varied from 93 to 188 stems in two-species stands, from 23 to 48 trees in multi-species stands with two main species, and from 21 to 46 trees in multi-species stands with no main species (Table 1).

According to McLachlan and Peel [35], a random variable X has a finite mixture distribution and that $f_X(\bullet)$ is a finite mixture probability density function (PDF) when:

$$f_X(x|\Psi) = \sum_{i=1}^k \pi_i f_i(x|\theta_i), \quad x \in \mathbf{X} \tag{1}$$

and

$$\Psi = \begin{pmatrix} \pi_1 & \dots & \pi_i & \dots & \pi_k \\ \theta_1^T & \dots & \theta_i^T & \dots & \theta_k^T \end{pmatrix} \tag{2}$$

π_i are called weights (fractions); $f_i(\bullet)$ are component densities; $i = 1, 2, \dots, k$; $0 \leq \pi_i \leq 1$; $\sum_{i=1}^k \pi_i = 1$; θ_i denotes the parameters of the $f_i(\bullet)$ distribution, Ψ is a complete parameter set for the overall distribution.

In this study, two theoretical distributions were taken into account. The Weibull distribution and the gamma distribution, as PDFs of two mixture distributions consisting of $k = 2$ individual PDF components. The number of components was assumed, based on the analysis of the problem and on the purpose of the research; two age cohorts determine two components in the mixture. The functions consisting of two Weibull or two gamma distributions can be written as:

$$f_{(Weib)X}(x|\Psi) = \pi_1 f_{(Weib)1}(x|\theta_1) + (1 - \pi_1) f_{(Weib)2}(x|\theta_2) \tag{3}$$

and

$$f_{(gam)X}(x|\Psi) = \pi_1 f_{(gam)1}(x|\theta_1) + (1 - \pi_1) f_{(gam)2}(x|\theta_2), \tag{4}$$

respectively, where $\theta_i = (\alpha_i, \beta_i, \gamma_i)$ or $\theta_i^* = (\mu_i, \sigma_i, \gamma_i)$. The vectors θ_i and θ_i^* serve to provide initial values based on the empirical data;

see e.g., Macdonald et al. [30]; α_i, β_i are the shape and the scale parameter, respectively; γ_i is the location parameter ($x \geq \gamma_i$); μ_i, σ_i are the mean and the standard deviation, respectively; $i = 1, 2$. Therefore, each of these mixture distributions is characterised by seven parameters: $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2$ (or $\mu_1, \sigma_1, \gamma_1, \mu_2, \sigma_2, \gamma_2$), and the parameter π_1 characterising the optimal mixture.

The three-parameter Weibull distribution has a PDF given by

$$f_{(Weib)X}(x|\alpha, \beta, \gamma) = \frac{\alpha}{\beta} \left(\frac{x - \gamma}{\beta} \right)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)^\alpha} \tag{5}$$

The three-parameter gamma distribution has a PDF given by

$$f_{(gam)X}(x|\alpha, \beta, \gamma) = \frac{(x - \gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x-\gamma}{\beta}} \tag{6}$$

where $\Gamma(\bullet)$ is the gamma function.

The maximum likelihood estimation (MLE) method is used. The log-likelihood function ($ll_1(\Psi)$) and the minus log-likelihood function ($ll_2(\Psi)$) are given by [10]:

$$ll_1(\Psi) = \sum_{j=1}^l n_j \log P_j(\Psi) \tag{7}$$

$$ll_2(\Psi) = -2 \sum_{j=1}^l n_j \log \left(\frac{P_j(\Psi)}{O_j} \right) \tag{8}$$

where $P_j(\Psi)$ is the theoretical probability that an individual belongs to the j th interval, $O_j = \frac{n_j}{N}$ denotes the observed relative frequency of the j th interval, and l is the number of intervals. The combination of the EM algorithm with the Newton-type method was used for minimising the $ll_2(\Psi)$ function (Eq. (8)) for estimating the parameters of two-component mixtures [10,36]. This procedure starts from an initial value (see also Eqs. (3) and (4)):

$$\Psi^0 = \begin{pmatrix} \pi_1^0 & \pi_2^0 \\ \theta_1^{0T} & \theta_2^{0T} \end{pmatrix} \tag{9}$$

or using initial values strictly connected with the empirical data; see e.g., Macdonald et al. [30]:

$$\Psi^0 = \begin{pmatrix} \pi_1^0 & \pi_2^0 \\ \theta_1^{*0T} & \theta_2^{*0T} \end{pmatrix} \tag{10}$$

Four methods for choosing initial values for the numerical procedure were analysed [53,54]:

Table 1
Basal area of species investigated and tree number per plot.

Statistic	Dominant species				Other species			All species	
	<i>sp1</i>		<i>sp2</i>		Species number	(m ² /ha)	(%)	(m ² /ha)	Tree number
	(m ² /ha)	(%)	(m ² /ha)	(%)					(N/plot)
<i>Two-species stands</i>									
Minimum	1.57	11.2	1.86	7.5	0	0.00	0.0	11.58	93
Mean	16.87	57.1	12.73	40.7	1.4	0.48	2.1	30.08	126.6
Maximum	27.99	91.4	31.99	88.7	2	1.57	7.9	46.12	188
SD	8.82		10.19			0.48		10.64	
<i>Multi-species stands with two main species</i>									
Minimum	0.86	2.8	3.92	23.0	4	2.31	7.2	14.77	23
Mean	5.27	22.3	13.64	51.4	6.9	6.00	26.2	24.91	33.7
Maximum	10.95	47.3	26.27	83.9	11	10.45	39.6	41.00	48
SD	3.34		8.25			2.21		7.55	
<i>Multi-species stands with no main species</i>									
Minimum	0.83	4.9	2.45	22.4	5	4.57	41.8	10.93	21
Mean	3.58	16.0	9.06	35.0	7.6	12.00	49.0	24.64	30.3
Maximum	7.21	35.8	21.52	46.6	10	22.39	67.1	46.19	46
SD	1.87		5.25			4.78		9.89	

1. $\min.k/\max.k$ [46,47]:

$$\Psi^0 = \begin{pmatrix} 0.5 & 0.5 \\ \theta_1^{*OT} & \theta_2^{*OT} \end{pmatrix}$$

with $\theta_1^0 = (\min.k, s)$ and $\theta_2^0 = (\max.k, s)$;

where $\min.k$, $\max.k$, and s are k -minimum, k -maximum, and standard deviation (SD) values of the DBH of all trees in the plot investigated, respectively; $\min.k$ is the k th smallest, and $\max.k$ is the k th largest DBH value in a data set (consists of all DBHs in the plot investigated); for each pair $\min.k < \max.k$.

In this study: $\min.k \equiv \min.1$ (\equiv minimum), $\min.5$, and $\min.10$; $\max.k \equiv \max.1$ (\equiv maximum), $\max.5$, and $\max.10$. The numerical procedure starts from $\min.1$ and $\max.1$; then from $\min.5$ and $\max.5$; finally, from $\min.10$ and $\max.10$. It starts 3 times, and then should be examine the results to see whether the same solution was obtained each time.

2. $0.5/1.5/\text{mean}$:

$$\Psi^0 = \begin{pmatrix} 0.5 & 0.5 \\ \theta_1^{*OT} & \theta_2^{*OT} \end{pmatrix}$$

with $\theta_1^0 = (0.5m, s)$ and $\theta_2^0 = (1.5m, s)$;

where m is the mean DBH of all trees in the plot investigated.

3. sp :

$$\Psi^0 = \begin{pmatrix} w_{sp1} & w_{sp2} \\ \theta_1^{*OT} & \theta_2^{*OT} \end{pmatrix}$$

with $\theta_1^0 = (m_{sp1}, s_{sp1})$ and $\theta_2^0 = (m_{sp2}, s_{sp2})$;

where the subscripts $sp1$ and $sp2$ indicate two dominant tree species on the plot investigated; w_{sp1} and w_{sp2} are the actual values of the weight (fraction) of trees belonging to the dominant tree species, respectively; m_{sp1} , m_{sp2} , s_{sp1} , and s_{sp2} are the actual values of the mean DBH and the SD of DBH for dominant tree species, respectively; $m_{sp1} \leq m_{sp2}$.

4. Multistart method [6]:

the grid points on the data space (10 points were employed) are given by:

$$u_1 \equiv \min, u_2 = u_1 + \frac{u_1 - u_{10}}{9}, \dots, u_9 = u_8 + \frac{u_1 - u_{10}}{9}, u_{10} \equiv \max$$

where \min , and \max are minimum, and maximum values of the DBH of all trees in the plot investigated, respectively.

Each subset of initial values is given by:

$$\psi_{c_1, c_2}^0 = \begin{pmatrix} 0.5 & 0.5 \\ \theta_{c_1}^{*OT} & \theta_{c_2}^{*OT} \end{pmatrix}$$

with: $(\theta_{c_1}^0 = (u_1, s), \theta_{c_2}^0 = (u_2, s)), \dots, (\theta_{c_1}^0 = (u_1, s), \theta_{c_2}^0 = (u_{10}, s)),$
 $(\theta_{c_1}^0 = (u_2, s), \theta_{c_2}^0 = (u_3, s)), \dots, (\theta_{c_1}^0 = (u_2, s), \theta_{c_2}^0 = (u_{10}, s)), \dots, (\theta_{c_1}^0 = (u_9, s),$
 $\theta_{c_2}^0 = (u_{10}, s)).$

A complete set of initial values ψ^0 consists of 45 ψ_{c_1, c_2}^0 subsets. The numerical procedure starts 45 times, and then should be examine the results to see whether the same solution was obtained each time.

The likelihood-ratio chi-square test was chosen to assess the goodness-of-fit of the models investigated [31,50]:

$$\chi^2 = -2 \sum_{j=1}^l n_j \log \left(\frac{\hat{n}_j}{n_j} \right) \quad (11)$$

where n_j and \hat{n}_j are the observed and predicted numbers of trees, respectively, in the j th DBH class in the plot; l is the number of DBH classes. The chi-square test has $(l - np - 1)$ degrees of freedom, where np is the number of parameters estimated. In addition to the likelihood-ratio chi-square test, four local error indexes (DIF.k+, DIF.k-; for $k = 1, 3$) were proposed:

$$\text{DIF.k}_+ = \frac{\max.k(n_j - \hat{n}_j)}{N} \quad (12)$$

$$\text{DIF.k}_- = \frac{\min.k(n_j - \hat{n}_j)}{N} \quad (13)$$

where $\max.k$ is the k th largest and $\min.k$ is the k th smallest differences between observed and predicted numbers of trees (the difference is in the j th DBH class) in the plot, respectively. N is the number of all trees in the plot. In this study: $\max.k \equiv \max.1$ (\equiv maximum), and $\max.3$, as well as $\min.k \equiv \min.1$ (\equiv minimum), and $\min.3$. The model with the indexes closest to zero was considered as the most suitable for DBH modelling.

The absolute errors (E , – calculated for each plot) and the mean absolute error (ME – calculated for each group of plots) were used to evaluate the errors for the predicted values of the mean DBH, and the SD of DBH of the two dominant species distinguished in the plots investigated. The errors were defined as:

$$E_{m_{sp1}} = |m_{sp1} - \hat{\mu}_1| \text{ and } E_{m_{sp2}} = |m_{sp2} - \hat{\mu}_2| \quad (14)$$

$$E_{s_{sp1}} = |s_{sp1} - \hat{\sigma}_1| \text{ and } E_{s_{sp2}} = |s_{sp2} - \hat{\sigma}_2| \quad (15)$$

$$ME = \frac{1}{Ng} \sum_{pl=1}^{Ng} E. \quad (16)$$

where m_{sp1} , m_{sp2} , s_{sp1} , s_{sp2} , and $\hat{\mu}_i$, $\hat{\sigma}_i$ are the actual and predicted (component) values of the mean DBH, and the SD of DBH of the two dominant species in the plot investigated, respectively; $i = 1, 2$; pl indicates the plot, $pl = 1, 2, \dots, Ng$; Ng is the number of plots in the group g ; $g = 1, 2, 3$. Small values of the errors (especially $E_{m_{sp1}}$ and $E_{m_{sp2}}$) permit the assumption that the DBH components may be associated with dominant species.

The kernel-type estimators are commonly used nonparametric estimators for density functions (e.g., [51,41]). Let x_1, \dots, x_n be sample points from an unknown density f . Then, its kernel estimate \hat{f} is:

$$\hat{f}(x|h) = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) \quad (17)$$

where $K(\bullet)$ is a kernel function, h is a bandwidth. In this study a Gaussian density as the kernel and a bandwidth $h = 2$ cm were used.

In order to verify the precision of the approximation of empirical DBH data using the two-component mixture distributions and the kernel density estimation, two statistics were employed:

$$B = \frac{1}{l} \sum_{j=1}^l (n_j - \hat{n}_j) \quad (18)$$

$$A = \frac{1}{l} \sum_{j=1}^l |n_j - \hat{n}_j| \quad (19)$$

where n_j and \hat{n}_j are the observed and predicted numbers of trees (for the mixture Weibull model, the mixture gamma model or the kernel density estimate), respectively, in the j th DBH class in the plot; l is the number of DBH classes.

The Fan's T test [13], in the form T_1 proposed by Pagan and Ullah [40], considers the difference between the two-component mixture distributions and the kernel density estimate. The T_n test statistic has a center term that may contribute to some finite sample bias [28]. To eliminate this effect a modified T_{1n} test statistic was used [40,28]. For a Gaussian density kernel T_{1n} has the following form [28]:

$$T_{1n} = \frac{1}{\sqrt{\sqrt{2}\hat{\sigma}}} n\sqrt{h} \left(J_n - \frac{1}{2\sqrt{\pi}nh} \right) \quad (20)$$

where $\hat{\sigma}$ is proportional to the integrated squared kernel density estimate; n is the sample size; h is a bandwidth; J_n is the integrated squared difference between the kernel density estimate and the two-component mixture distributions – for the mixture Weibull model or the mixture gamma model (see Pagan and Ullah [40] as well as Li et al. [28] for details). Under the null hypothesis of adequation, this test is asymptotically $N(0, 1)$.

The calculations were carried out using the *mixdist* package of R [31,48] and Mathematica 8 (Wolfram).

3. Results

In stands from the Świętokrzyskie Mountains fir *Abies alba* Mill. and beech *Fagus sylvatica* L. prevailed. In these stands fir and beech formed two dominant species. The stands from the southern Appalachians were composed of eastern white pine *Pinus strobus* L., loblolly pine *P. taeda* L., Virginia pine *P. virginiana* Mill., eastern hemlock *Tsuga canadensis* (L.) Carr., southern red oak *Quercus falcata* Michx., white oak *Q. alba* L., northern red oak *Q. rubra* L., scarlet oak *Q. coccinea* Muenchh., chestnut oak *Q. prinus* L., mountain magnolia *Magnolia fraseri* Walt., sweetgum *Liquidambar styraciflua* L., red maple *Acer rubrum* L., yellow-poplar *Liriodendron tulipifera* L., pignut hickory *Carya glabra* (Mill.) Sweet, black cherry *Prunus serotina* Ehrh., sweet birch *Betula lenta* L., sourwood *Oxydendrum arboreum* (L.) DC., and black locust *Robinia pseudoacacia* L. In the southern Appalachians, pines and oaks dominated among the most prevalent species (they accounted for more than 35% of the basal area). The basal area percentage of dominant species *sp1* (most likely associated with the young age cohort) varied from 11.2% to 91.4% in two-species stands, from 2.8% to 47.3% in multi-species stands with two main species, and from 4.9% to 35.8% in multi-species stands with no main species (Table 1). The basal area percentage of dominant species *sp2* (initially associated with the old age cohort) varied from 7.5% to 88.7% in two-species stands, from

23.0% to 83.9% in multi-species stands with two main species, and from 22.4% to 46.6% in multi-species stands with no main species (Table 1). The other species in individual tree groups reached a maximum of 7.9%, 39.6%, and 67.1%, respectively (Table 1). The total basal area varied from 11.58 m²/ha to 46.12 m²/ha in two-species stands, from 14.77 m²/ha to 41.00 m²/ha in multi-species stands with two main species, and from 10.93 m²/ha to 46.19 m²/ha in multi-species stands with no main species (Table 1).

The mean DBH was lower for dominant species *sp1* than for dominant species *sp2* in all of the groups investigated (Table 2). The species most likely associated with the young age cohort were characterised by lower diameter variation in comparison to the species initially associated with the old age cohort. The mean SD of DBH ranged from 5.0 cm to 13.0 cm, and from 11.8 cm to 13.9 cm in the young and old cohorts, respectively (Table 2). Tree DBH distributions displayed positive skewness, i.e., asymmetry towards positive values. The greatest asymmetry occurred in DBH distributions in plots representing the multi-species stands with no dominant species (mean skewness was 1.6963); plots representing the two-species stands were less asymmetrical (mean skewness was 1.2034) (Table 2). The DBHs measured ranged from 93 to 188 in two-species stands in the Świętokrzyskie Mountains and from 21 to 48 in multi-species stands in the southern Appalachians (Table 2).

To find parameters of two-component mixture models, first the weights, the means, and the SDs of the models were estimated, and next, the appropriate shapes and scales were calculated. The average values of weights (π_i) showed the greater proportion of the 1st component in the models (Table 3). Greater differences between component distributions for the 2nd component of mixture models were observed when analysing the shape of PDF function for particular groups (Table 3).

In the two-species stands, the consistency of empirical data with the appropriate theoretical distribution was achieved using the mixture Weibull model for 6 plots, and using the mixture gamma model for the remaining 7 plots (χ^2 test, $P > 0.05$; Table 4). For multi-species stands dominated by two species, the empirical data were in accordance with mixture Weibull and gamma models for 9 of the 10 plots (χ^2 test, $P > 0.05$; Table 4). For multi-species stands with no main species, the consistency of empirical data with the appropriate theoretical distribution was achieved using the mixture Weibull model for 7 plots, and using the mixture gamma model for 8 plots (χ^2 test, $P > 0.05$; Table 4). The best fit for the two analysed models was achieved for multi-species stands with no

Table 2
Statistics of the tree DBHs.

Statistic	Dominant species <i>sp1</i>		Dominant species <i>sp2</i>		All species					
	Mean DBH (cm)	SD of DBH (cm)	Mean DBH (cm)	SD of DBH (cm)	Mean DBH (cm)	SD of DBH (cm)	Max DBH (cm)	Skewness	Kurtosis	DBHs measured (N/plot)
<i>Two-species stands</i>										
Minimum	12.7	4.8	15.1	5.6	14.7	8.8	45	0.4148	-0.9760	93
Mean		13.0		13.9		14.3	67	1.2034	1.5140	126.6
Maximum	27.7	17.5	37.0	29.7	30.6	21.3	90	2.7717	9.5660	188
SD							13	0.7449	3.1130	
<i>Multi-species stands with two main species</i>										
Minimum	15.1	2.0	23.7	5.3	17.8	5.4	32	0.1449	-1.1600	23
Mean		7.4		11.8		9.8	51	1.3432	2.4175	33.7
Maximum	26.6	25.0	40.8	29.0	28.8	13.8	71	3.4107	13.7273	48
SD							13	0.9386	4.6326	
<i>Multi-species stands with no main species</i>										
Minimum	15.3	2.1	18.8	3.8	19.0	7.7	45	0.7588	-0.8065	21
Mean		5.0		12.4		11.1	56	1.6963	3.3679	30.3
Maximum	32.2	9.4	63.5	20.8	29.1	17.1	78	2.8710	10.5005	46
SD							12	0.6649	3.4436	

Table 3
Average values of the parameters and statistics of the two-component mixture models.

Statistic	1st Component					2nd Component				
	Weight	Shape	Scale	Mean	SD	Weight	Shape	Scale	Mean	SD
<i>Mixture Weibull model</i>										
Two-species stands										
Minimum	0.175	0.694	4.948	4.38	2.04	0.032	2.053	21.503	19.05	1.34
Mean	0.678	1.261	9.816			0.322	11.358	40.899		
Maximum	0.968	2.275	18.656	17.59	16.87	0.825	61.397	78.517	76.57	14.60
SD	0.241	0.456	4.585			0.241	16.802	17.268		
Multi-species stands with two main species										
Minimum	0.279	0.911	2.622	2.54	0.92	0.117	1.070	12.086	10.74	3.49
Mean	0.589	2.054	5.937			0.411	3.508	23.936		
Maximum	0.884	5.087	13.924	12.68	6.70	0.721	7.991	50.298	46.98	12.73
SD	0.208	1.391	4.167			0.208	2.134	11.035		
Multi-species stands with no main species										
Minimum	0.497	0.952	3.095	2.79	1.85	0.067	0.545	4.607	7.95	2.86
Mean	0.765	1.488	6.643			0.235	6.539	28.857		
Maximum	0.933	2.545	10.369	9.84	9.57	0.503	15.679	37.827	36.49	15.80
SD	0.160	0.468	2.411			0.160	5.343	10.213		
<i>Mixture gamma model</i>										
Two-species stands										
Minimum	0.283	0.533	1.135	4.73	2.32	0.031	4.586	0.188	23.99	3.15
Mean	0.719	1.546	9.252			0.281	86.637	2.410		
Maximum	0.969	4.164	18.935	17.44	15.10	0.717	398.544	6.227	76.70	14.46
SD	0.209	0.977	5.250			0.209	136.525	2.089		
Multi-species stands with two main species										
Minimum	0.275	0.754	0.671	2.66	1.71	0.099	1.151	0.294	12.36	2.65
Mean	0.646	3.271	3.237			0.354	19.941	3.158		
Maximum	0.901	13.863	7.106	13.08	7.83	0.725	81.387	11.571	48.43	12.42
SD	0.189	3.873	2.061			0.189	22.918	3.306		
Multi-species stands with no main species										
Minimum	0.520	0.889	1.130	2.95	2.04	0.063	0.326	0.168	7.42	2.48
Mean	0.785	2.057	5.224			0.215	57.820	4.013		
Maximum	0.937	6.362	13.118	15.30	14.17	0.480	218.334	22.775	36.63	13.55
SD	0.146	1.581	3.853			0.146	71.693	6.841		

main species (Table 4). The average values $P(\chi^2 \text{ test})$ show that empirical DBH data was best approximated by the mixture Weibull model, however, the mixture gamma model approximated the data only slightly worse and converged more consistently (Table 4).

The values of local error indexes calculated for the analysed models varied, for $DIF.k_+$, from 0.007 to 0.152 for the mixture Weibull model and from 0.010 to 0.142 for the mixture gamma model, while $DIF.k_-$, ranged from -0.162 to -0.010 for the mixture Weibull model and from -0.158 to -0.010 for the mixture gamma model (Table 4). The greatest average maximal differences between empirical data and the analysed models were shown for the multi-species stands, and the smallest for two-species stands. The application of the mixture gamma model instead of the mixture Weibull model allows one to obtain slightly smaller average maximal values between empirical data and theoretical distributions (Table 4).

The analyses presented here have shown that in two-cohort, two-storied, multi-species stands with two main species and with no main species from the southern Appalachians, two-component mixtures of the same distributions of the Weibull and gamma distribution are appropriate models for the DBHs. In the case of two-cohort, partly multilayered stands from the Świętokrzyski National Park the analysed models appear to be less useful. Additionally the mixture Weibull model and the mixture gamma model approximated the empirical DBH distributions with similar precision.

The indexes $DIF.k_+$ and $DIF.k_-$ varied, for large samples (in the Świętokrzyskie Mountains), from 0.010 to 0.121 and from -0.092 to -0.010 , respectively, as well as for small samples (in the southern Appalachians), from 0.007 to 0.152 and from -0.162 to -0.010 , respectively (Table 4). In this study the goodness-of-fit was not influenced by the number of DBHs measured.

Theoretically, in mixed-species, two-cohort stands, the two-component DBH structure can be associated with: (1) age cohorts and tree species – the first dominant species (in this paper $sp1$) dominates a young age cohort, and the second one (in this paper $sp2$) dominates an old age cohort; or (2) only age cohorts – dominant species built both generations; a given species occurs in both young and old age cohorts. The mean absolute error of mean DBH ranged from 2.4 cm to 5.1 cm for the mixture Weibull model and from 2.6 cm to 4.5 cm for the mixture gamma model for $sp1$, while it ranged from 5.5 cm to 18.1 cm for the mixture Weibull model and from 5.8 cm to 19.5 cm for the mixture gamma model for $sp2$ (Table 5). The maximum absolute error for mean DBH was for dominant species $sp1$ smaller than 15 cm (mixture Weibull model) and 12 cm (mixture gamma model), and for dominant species $sp2$ smaller than 47 cm (for both tested models) (Table 5). The maximum absolute error for SD of DBH was for dominant species $sp1$ smaller than 22 cm (for both tested models), and for dominant species $sp2$ smaller than 26 cm (also for both tested models) (Table 5).

The greatest errors of mean DBH were shown for two-species stands from the Świętokrzyski National Park. The mean absolute error of mean DBH for dominant species $sp1$ was 5.1 cm for the mixture Weibull model and 4.5 cm for the mixture gamma model, while that for dominant species $sp2$ was 18.1 for the mixture Weibull model and 19.5 cm for the mixture gamma model (Table 5). When analysing SD of DBH for two-species stands from the Świętokrzyski National Park, we obtained the following values: the mean absolute error for SD of DBH was 4.2 cm and 3.5 cm for dominant species $sp1$ for the mixture Weibull model and the mixture gamma model, respectively, and 5.9 cm and 6.0 cm for dominant species $sp2$ for the mixture Weibull model and the mixture gamma model, respectively (Table 5).

Table 4
Likelihood-ratio χ^2 test and local error indexes of the two-component mixture models.

Plot	Mixture Weibull model				Mixture gamma model					
	χ^2 test	Local error index			χ^2 test	Local error index				
		P value	DIF.1+	DIF.3+		DIF.1-	DIF.3-	P value	DIF.1+	DIF.3+
<i>Two-species stands</i>										
ŠPN21	0.4176	0.024	0.017	-0.046	-0.021	0.3094	0.026	0.023	-0.043	-0.024
ŠPN22	0.9255	0.023	0.014	-0.034	-0.017	0.9023	0.020	0.013	-0.034	-0.014
ŠPN23	0.0027	0.058	0.020	-0.058	-0.023	0.0017	0.059	0.018	-0.060	-0.021
ŠPN24	0.4614	0.017	0.016	-0.024	-0.017	0.3884	0.017	0.015	-0.026	-0.019
ŠPN25	0.0391	0.046	0.020	-0.040	-0.019	0.0796	0.047	0.015	-0.037	-0.017
ŠPN26	0.0594	0.059	0.031	-0.050	-0.022	0.0784	0.046	0.029	-0.051	-0.022
ŠPN27	0.0171	0.047	0.029	-0.040	-0.022	0.0093	0.044	0.028	-0.042	-0.021
ŠPN28	<0.001	0.121	0.031	-0.089	-0.038	<0.001	0.119	0.032	-0.092	-0.037
ŠPN29	0.7915	0.016	0.014	-0.015	-0.011	0.7591	0.018	0.014	-0.023	-0.011
ŠPN30	0.3849	0.047	0.011	-0.047	-0.013	0.3344	0.050	0.010	-0.053	-0.010
ŠPN31	0.0065	0.028	0.020	-0.039	-0.022	0.0106	0.027	0.020	-0.035	-0.017
Mean	0.2823	0.0442	0.0203	-0.0438	-0.0205	0.2612	0.0430	0.0197	-0.0451	-0.0194
<i>Multi-species stands with two main species</i>										
3028	0.2266	0.042	0.037	-0.046	-0.031	0.1946	0.044	0.039	-0.046	-0.035
37022	0.1724	0.038	0.029	-0.051	-0.029	0.1791	0.043	0.032	-0.052	-0.027
71025	0.2936	0.068	0.029	-0.029	-0.019	0.2926	0.068	0.029	-0.029	-0.018
57049	0.0614	0.088	0.027	-0.057	-0.028	0.1045	0.071	0.032	-0.053	-0.029
11025	0.2905	0.125	0.029	-0.066	-0.025	0.2400	0.128	0.029	-0.059	-0.023
23057	0.1654	0.071	0.033	-0.052	-0.041	0.1244	0.077	0.050	-0.050	-0.043
75028	0.9595	0.050	0.020	-0.047	-0.019	0.9595	0.049	0.021	-0.045	-0.015
115045	0.4625	0.091	0.057	-0.042	-0.028	0.4921	0.093	0.042	-0.042	-0.023
23046	0.0255	0.066	0.044	-0.162	-0.033	0.0273	0.068	0.048	-0.158	-0.028
193028	0.1956	0.066	0.007	-0.040	-0.014	0.1542	0.068	0.014	-0.037	-0.019
Mean	0.2853	0.0705	0.0312	-0.0592	-0.0267	0.2768	0.0709	0.0336	-0.0571	-0.0260
<i>Multi-species stands with no main species</i>										
75045	0.1111	0.090	0.064	-0.129	-0.031	0.0861	0.098	0.064	-0.130	-0.030
183084	0.5500	0.036	0.023	-0.042	-0.022	0.5478	0.035	0.022	-0.043	-0.022
111039	^a	-	-	-	-	0.8411	0.079	0.045	-0.043	-0.032
99043	0.6513	0.042	0.035	-0.047	-0.022	0.6325	0.049	0.034	-0.048	-0.023
71013	0.6500	0.046	0.017	-0.050	-0.010	0.6350	0.042	0.016	-0.049	-0.010
113079	0.6767	0.072	0.017	-0.036	-0.018	0.7086	0.066	0.018	-0.034	-0.021
199016	0.4292	0.043	0.028	-0.069	-0.045	0.4676	0.043	0.027	-0.067	-0.037
27013	0.0097	0.145	0.028	-0.038	-0.028	0.0092	0.142	0.030	-0.042	-0.028
39065	0.4305	0.152	0.046	-0.093	-0.034	0.4320	0.141	0.046	-0.097	-0.023
Mean	0.4386	0.0783	0.0323	-0.0630	-0.0263	0.4844	0.0772	0.0336	-0.0614	-0.0251

^a Lack of convergence of the estimation process; in the paper: DIF.k+, DIF.k- (for k = 1,3) are the local error indexes (see Eqs. (12) and (13)).

In addition, one may observe that there are only minor differences between the Weibull and gamma mixture models, and either of them may be used for analysing the relationships between the two-component DBH structure and species composition of stands.

The error values led to two conclusions. First, in the two-cohort, two-storied, multi-species stands, with both two main species and with no main species, from the southern Appalachians, the 1st DBH component of mixture models was associated with dominant species *sp1* and, to a lesser degree, the 2nd DBH component was associated with dominant species *sp2* (Figs. 1 and 2). Secondly, in most of the two-cohort, partly multilayered stands from the Świętokrzyski National Park, the two-component DBH structure was associated with only age cohorts (Fig. 3). In the southern Appalachians the young age cohorts were composed of dominant species *sp1* belonging to the group of both shade-intolerant species, such as sweetgum and black cherry, and shade-tolerant species such as eastern hemlock and sourwood. Similarly, the old age cohorts were composed of dominant species *sp2* from the group of shade-intolerant species (e.g., loblolly pine and yellow-poplar) as well as shade-tolerant species (e.g., red maple and sourwood). Generally, in a lower layer in multi-species stands with two main species, and with no main species, there occurred dominant species *sp1* of different shade tolerances. In the upper canopy layer of multi-species stands with two main species there prevailed dominant species *sp2* representing shade-intolerant taxa, while in multi-spe-

cies stands with no main species there occurred mostly dominant species *sp2* of average tolerance to shading (e.g., eastern white pine and chestnut oak). In the Świętokrzyski National Park beech and fir occurred together in most DBH classes. The young and old age cohorts were characterised by a similar species composition; beech and fir trees were growing in the lower and upper stand layers.

Estimation of the parameters of mixture models requires determining the global maximum of the likelihood function $ll_1(\Psi)$ (which is tantamount to the determination of global minimum of the likelihood function $ll_2(\Psi)$; see Eqs. (7) and (8)). The effective determination of global extreme values depends, among other things, on the choice of initial values. Here, we compared four methods to choose the initial values; for each of them we compared values at the solution points for the minus log-likelihood function (Table 6). Of the four methods, the best results were obtained from the multistart method (Table 6). The application of this method allowed us to determine the global minimum for 29 plots with the mixture Weibull model and for 30 plots with the mixture gamma model (Table 6). Of the remaining three methods the best was min.k/max.k (which enabled the determination of the global minimum for 25 plots for each tested model), followed by the 0.5/1.5/mean and sp methods (which, with similar effectiveness, enabled the determination of the global minimum for 19 and 16 plots, respectively) (Table 6). The local minimum or saddle point instead of a global minimum was found for 9 and 16 cases using

Table 5
The errors for the predicted values of the tree DBH statistics for two dominant species.

Statistic	Dominant species <i>sp1</i>		Dominant species <i>sp2</i>	
	Error for mean DBH (cm)	Error for SD of DBH (cm)	Error for mean DBH (cm)	Error for SD of DBH (cm)
<i>Mixture Weibull model</i>				
Two-species stands				
Minimum	0.1	0.8	1.5	0.8
Quartile _{0.25}	2.3	1.8	6.8	2.1
Mean	5.1	4.2	18.1	5.9
Quartile _{0.75}	7.4	6.1	26.0	6.8
Maximum	14.6	8.9	46.5	25.2
Multi-species stands with two main species				
Minimum	0.3	0.8	0.4	0.1
Quartile _{0.25}	1.2	1.9	1.5	1.5
Mean	2.7	5.0	5.5	4.6
Quartile _{0.75}	3.2	4.0	5.3	4.8
Maximum	8.6	21.7	22.0	16.3
Multi-species stands with no main species				
Minimum	1.1	0.9	0.6	0.5
Quartile _{0.25}	1.3	1.0	3.2	4.1
Mean	2.4	2.4	9.0	6.5
Quartile _{0.75}	3.2	2.6	12.3	10.2
Maximum	5.1	7.5	25.2	12.0
<i>Mixture gamma model</i>				
Two-species stands				
Minimum	0.2	0.2	1.7	0.4
Quartile _{0.25}	2.2	2.0	9.0	2.6
Mean	4.5	3.5	19.5	6.0
Quartile _{0.75}	6.5	4.6	27.1	6.5
Maximum	11.7	8.6	46.6	25.8
Multi-species stands with two main species				
Minimum	0.1	0.4	0.3	1.0
Quartile _{0.25}	1.4	2.5	0.8	2.2
Mean	2.6	4.8	5.8	5.4
Quartile _{0.75}	3.1	3.7	6.9	6.4
Maximum	7.7	21.5	23.5	16.9
Multi-species stands with no main species				
Minimum	0.7	0.9	0.1	0.5
Quartile _{0.25}	1.3	1.2	4.7	5.7
Mean	2.6	3.4	12.2	7.7
Quartile _{0.75}	3.4	3.6	17.6	10.4
Maximum	5.0	10.8	31.2	15.9

In the paper: E_i are the absolute errors (see Eqs. (14) and (15)), mean absolute error (ME) is the mean (see Eq. (16)).

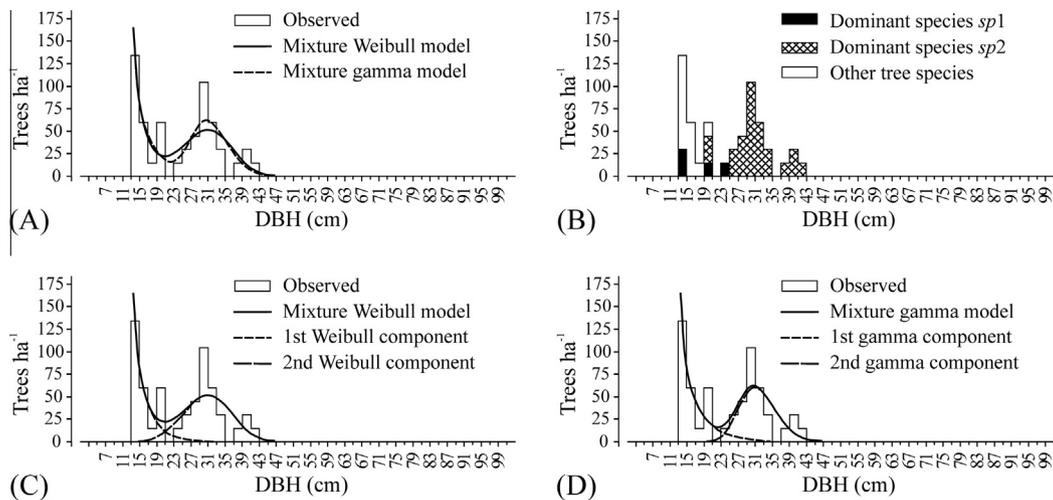


Fig. 1. Multi-species stand with two main species (plot 57049); the 1st DBH component of the mixture models was associated with dominant species *sp1* (black cherry) and the 2nd DBH component was associated with dominant species *sp2* (loblolly pine). (A) Mixture models, (B) DBH species structure, (C) mixture Weibull model, (D) mixture gamma model.

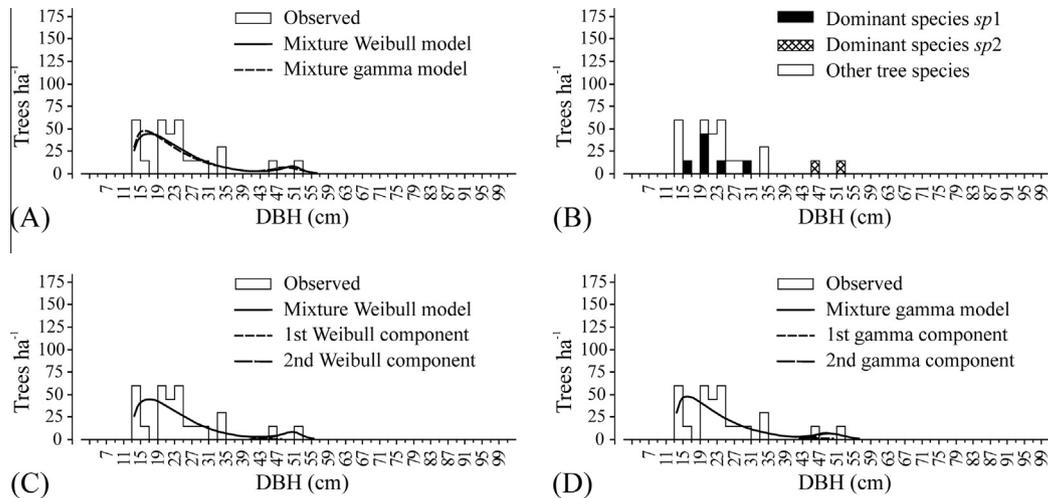


Fig. 2. Multi-species stand with no main species (plot 75045); the 1st DBH component of the mixture models was associated with dominant species *sp1* (red maple) and the 2nd DBH component was associated with dominant species *sp2* (mountain magnolia). (A) Mixture models, (B) DBH species structure, (C) mixture Weibull model, (D) mixture gamma model.

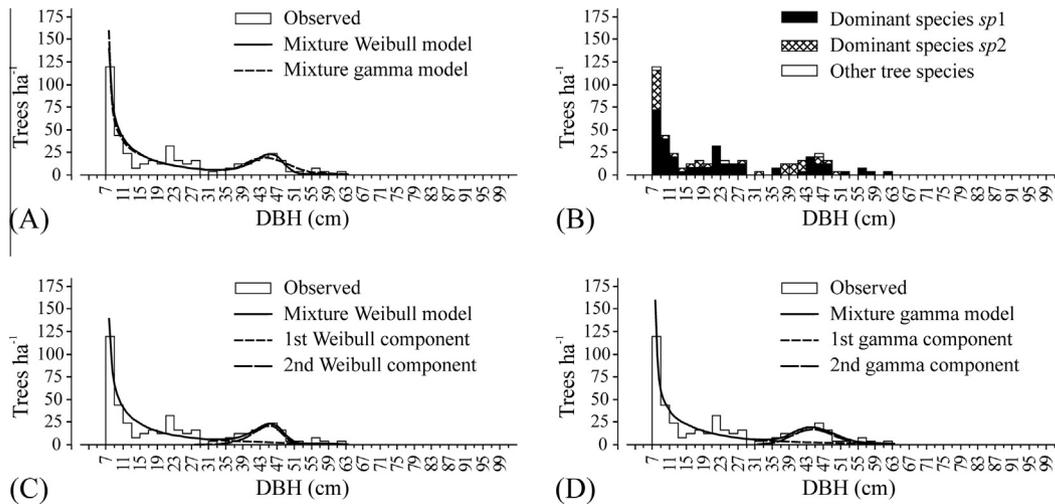


Fig. 3. Two-species stands (plot ŠPN25); the DBH structure was associated only with age cohorts, beech (*sp1*) and fir (*sp2*) trees occurred in the young as well as in the old age cohort. (A) Mixture models, (B) DBH species structure, (C) mixture Weibull model, (D) mixture gamma model.

the mixture Weibull model and mixture gamma model, respectively, while lack of convergence of the estimation process occurred for mixture Weibull model and mixture gamma model in 19 and 17 cases, respectively (Table 6). The convergence was influenced by the sample size, but there are only minor differences between small and large samples – more problems occurred in the case of plots from the southern Appalachians, which were characterised by small data sets (Table 6). The results showed that multistart and $\min.k/\max.k$ methods are most useful for choosing initial values for the estimation of parameters for Weibull and gamma models.

The values of the B statistic are lowest for the kernel density estimator (they vary from 0.0004 to 0.0166), which indicates that the kernel density estimate has the lowest bias (Table 7). The bias is higher with mixture distribution approximation; the values of the B statistic take on values from 0.0004 to 0.1023 when the Weibull model was used and from 0.0003 to 0.166 for the gamma model (Table 7). The values of the A statistic are a measure of the flexibility of the analysed distributions. The lowest values of the A statistic were found for the Weibull and the gamma mixture

distributions (from 0.3040 to 1.9546; mean is 0.8148), and the highest ones when the kernel density estimator was used (from 0.2629 to 1.9883; mean is 0.9166) (Table 7).

Significant differences between the two-component mixture distributions and the kernel density estimator were obtained for 9 plots (Fan's T_1 test, $P < 0.05$; Table 7). For these 9 plots the values of the A statistic were higher when the kernel density estimator was used (Table 7). In this study, the Weibull and the gamma mixture distributions approximated the empirical DBH data slightly better than the kernel density estimator (Table 7; Figs. 4 and 5).

4. Discussion

In contrast to simple probability models, mixture models are much more universal and significantly more flexible for the approximation of multimodal empirical data. Two-component mixture models appear to be useful for both fitting the empirical data (e.g., [73,29,72,24]) and differentiating subpopulations (e.g., [69,71,46,47]). When DBH distributions of two age cohorts overlap

Table 6

Values at the solution points for the minus log-likelihood function for the two-component mixture models.

Method to choose initial values Plot	Mixture Weibull model				Mixture gamma model			
	min.k/max.k	0.5/1.5/mean	sp	multistart	min.k/max.k	0.5/1.5/mean	sp	multistart
<i>Two-species stands</i>								
ŠPN21	38.14	54.52	^a	38.14	40.74	54.24	54.24	40.74
ŠPN22	11.72	11.72	11.72	11.72	12.38	12.38	12.38	12.38
ŠPN23	42.13	42.13	42.13	42.13	43.61	43.61	43.61	43.61
ŠPN24	33.12	33.12	33.12	33.12	34.67	34.67	34.67	34.67
ŠPN25	36.23	36.23	36.23	36.23	33.09	33.09	33.09	33.09
ŠPN26	40.52	40.52	40.52	40.52	39.16	39.16	39.16	39.16
ŠPN27	42.21	42.21	42.21	42.21	44.61	44.61	44.61	44.61
ŠPN28	69.78	69.78	–	69.78	70.01	–	–	70.01
ŠPN29	10.44	10.44	–	10.44	10.91	12.67	–	10.91
ŠPN30	22.26	29.90	29.90	22.26	23.18	28.60	28.60	23.18
ŠPN31	67.51	–	53.95	53.95	51.93	69.80	51.93	51.93
<i>Multi-species stands with two main species</i>								
3028	11.77	11.77	11.77	11.77	12.34	12.34	12.34	12.34
37022	17.62	17.62	17.62	17.62	17.46	17.46	17.46	17.46
71025	19.63	19.77	19.63	19.63	19.65	19.69	19.65	19.65
57049	17.64	17.64	17.64	17.64	15.84	15.84	15.84	15.84
11025	26.22	26.22	26.22	26.22	28.96	28.96	28.96	27.38
23057	17.80	17.80	17.80	17.80	18.96	–	–	18.96
75028	14.13	14.31	–	14.13	–	14.13	–	14.13
115045	22.97	22.97	22.97	22.97	22.47	22.47	22.47	22.47
23046	–	–	–	20.42	–	–	–	20.21
193028	7.35	–	–	7.35	8.04	8.04	8.04	8.04
<i>Multi-species stands with no main species</i>								
75045	–	–	21.88	21.88	–	–	22.91	22.91
183084	11.73	11.73	11.73	11.73	11.76	11.76	11.76	11.76
111039	–	–	–	–	20.61	–	–	20.61
99043	24.57	24.57	24.57	24.57	24.92	24.92	27.57	24.92
71013	11.46	12.39	–	11.46	11.64	12.40	12.40	11.64
113079	10.18	10.18	10.18	9.30	8.93	10.18	10.18	8.93
199016	16.34	16.34	16.34	16.34	15.79	20.52	20.52	15.79
27013	32.09	32.09	32.09	32.09	32.29	32.29	32.29	32.29
39065	14.26	–	–	14.26	–	–	–	14.24

^a Lack of convergence of the estimation process; bold numbers indicate that the estimation process stopped in a good status; in the paper: $l_2(\psi)$ is the minus log-likelihood function (see Eq. (8)).

strongly, the resulting stand structures have reverse J and rotated-sigmoidal shapes. In those cases, the application of single, flexible distributions such as the Burr Type III distribution [17], is usually a better solution than the use of mixture models.

Empirical DBH distributions are often multimodal, and this multimodality can be random, local (due to abrupt, periodical changes in the growth dynamics of a group of trees, or small-scale disturbances) or general (connected with the existence of subpopulations in a stand). For small data sets (where there are less than 80 trees) there was apparent random, local multimodality, often in the form of several local extremes, while for large samples (for a few hundred trees) there were usually distinct (mostly two) local maxima connected with the existence of subpopulations. That is why, when determining the number of mixture model components, one should first make use of additional information about the possible existence of subpopulations in a stand, and then, if necessary, especially for small samples, apply suitable statistical tests (e.g., [35]). Summing up, the procedure one should follow depends on what is needed in the research. When one wants to present, as precisely as possible, DBH distribution of a specific data set, then one can fit the number of mixture model components to the number of random, local extremes. When the purpose of the study is to make certain generalisations, then one should ignore any random, local multimodality and focus on the separation of local maxima possibly connected with the existence of subpopulations. In mixed-species, two-cohort, two-storied and partly multilayered stands, subpopulations which can shape DBH distributions are for example particular tree species (especially dominant species),

age cohorts or stand layers. These elements are often interconnected, e.g., two-cohort or two-storied stands can be two-species stands and then it is possible that particular age cohorts or stand layers are formed by other tree species. Relationships between the two-component DBH structure, age cohorts and species composition of stands depend, among others, on forest dynamics, disturbances, habitat, type of regeneration and light requirements of particular species.

Disturbances are the basic factors determining spatio-temporal dynamics in forest stands (e.g., [18,42]). Most studies have centred on the extreme ends of the disturbance spectrum (e.g., [37,23]). Less attention has been focused on the intermediate-scale disturbances between these extremes (0.04–0.5 ha in extent), which may play a larger role in driving the dynamics of forests than previously anticipated [44]. Relatively discrete events in time which disrupt forest ecosystems create a continuous gradient of disturbance severity ranging from fine-scale to coarse-scale damage. Fine-scale disturbances (usually < 0.04 ha in extent) eliminate individual trees locally, while coarse-scale disturbances (usually > 0.5 ha in extent) cause catastrophic stand break-ups (e.g., [8,59]). Models based on birth, growth and death of patches analyse above of all the effect of disturbances of different size on forest dynamics (e.g., [52,26]). These models assume that there is a competitive gradient from shade-intolerant, fast-growing, pioneer species to shade-tolerant, slower-growing, advanced successional species. Shade-intolerant species are recruited in pulses, under the conditions created by perturbation, whereas shade-tolerant species are recruited more continuously (e.g., [39]).

Table 7
The goodness-of-fit statistics and the modified Fan's T_1 test for the two-component mixture distributions and the kernel density estimation.

Plot	Statistics						T_1 test			
	Mixture Weibull model		Mixture gamma model		Kernel model		Mixture Weibull model		Mixture gamma model	
	B	A	B	A	B	A	T_{1n}	P value	T_{1n}	P value
<i>Two-species stands</i>										
ŠPN21	0.0004	1.2183	0.0003	1.2622	0.0012	1.2777	-1.638	0.9493	-0.942	0.8269
ŠPN22	0.0127	0.8775	0.0235	0.8907	0.0058	1.1608	1.409	0.0794	1.502	0.0666
ŠPN23	0.0139	1.9503	0.0262	1.9546	0.0028	1.9883	2.568	0.0051	2.816	0.0024
ŠPN24	0.0134	0.7966	0.0243	0.7889	0.0019	0.8202	-1.975	0.9758	-1.848	0.9677
ŠPN25	0.1023	1.1748	0.1166	1.1910	0.0166	1.4404	13.965	<0.001	16.839	<0.001
ŠPN26	0.0072	1.2526	0.0137	1.1934	0.0004	1.1482	0.385	0.3500	1.555	0.0599
ŠPN27	0.0232	1.6790	0.0343	1.7211	0.0017	1.6597	-0.830	0.7967	-0.844	0.8007
ŠPN28	0.0087	1.6416	0.0132	1.6925	0.0021	1.7674	-1.034	0.8494	-0.951	0.8293
ŠPN29	0.0102	0.6757	0.0113	0.7518	0.0019	1.0633	-0.281	0.6106	-0.349	0.6364
ŠPN30	0.0048	0.8532	0.0072	0.8651	0.0126	1.2333	1.732	0.0416	2.641	0.0041
ŠPN31	0.0146	1.3949	0.0180	1.3334	0.0040	0.9916	1.421	0.0777	0.910	0.1814
Mean	0.0192	1.2286	0.0262	1.2404	0.0046	1.3228	1.429	0.4305	1.939	0.3978
<i>Multi-species stands with two main species</i>										
3028	0.0026	0.5695	0.0051	0.6312	0.0031	0.7989	0.692	0.2446	-0.123	0.5489
37022	0.0101	0.7148	0.0167	0.7244	0.0063	0.9509	1.932	0.0267	2.071	0.0192
71025	0.0110	0.4167	0.0108	0.4090	0.0078	0.8183	7.254	<0.001	7.590	<0.001
57049	0.0124	0.8876	0.0223	0.8009	0.0059	1.1328	3.286	0.0005	2.848	0.0022
11025	0.0018	0.6374	0.0048	0.6617	0.0005	0.5546	-1.248	0.8940	-1.149	0.8747
23057	0.0083	0.5204	0.0062	0.6184	0.0007	0.7441	5.336	<0.001	1.535	0.0624
75028	0.0043	0.3040	0.0040	0.3061	0.0012	0.2629	-1.613	0.9466	-1.658	0.9513
115045	0.0027	0.4482	0.0025	0.4240	0.0016	0.4006	-0.336	0.6314	0.060	0.4759
23046	0.0080	0.8507	0.0152	0.8498	0.0050	1.0175	0.640	0.2612	1.298	0.0971
193028	0.0010	0.6086	0.0036	0.6860	0.0082	1.0477	1.896	0.0289	1.175	0.1199
Mean	0.0062	0.5958	0.0091	0.6112	0.0040	0.7728	1.784	0.3034	1.365	0.3152
<i>Multi-species stands with no main species</i>										
75045	0.0006	0.6695	0.0029	0.6883	0.0024	0.5907	-0.210	0.5833	0.408	0.3416
183084	0.0078	0.4151	0.0104	0.4175	0.0044	0.6448	1.137	0.1278	1.179	0.1191
111039	^a	-	0.0040	0.3974	0.0006	0.3763	-	-	-1.739	0.9590
99043	0.0034	0.4668	0.0042	0.4742	0.0017	0.4020	-0.900	0.8160	-0.795	0.7867
71013	0.0022	0.5338	0.0027	0.4898	0.0053	0.7693	-0.330	0.6294	-0.244	0.5964
113079	0.0006	0.5687	0.0012	0.5428	0.0045	0.6612	-0.826	0.7955	-0.440	0.6701
199016	0.0080	0.4774	0.0098	0.4639	0.0019	0.4730	-0.861	0.8055	-0.837	0.7987
27013	0.0082	0.6619	0.0124	0.6848	0.0051	0.8212	1.824	0.0341	2.637	0.0042
39065	0.0296	0.5012	0.0514	0.5465	0.0028	0.5472	-0.464	0.6785	0.064	0.4744
Mean	0.0075	0.5368	0.0110	0.5228	0.0032	0.5873	-0.079	0.5588	0.026	0.5278

^a Lack of convergence of the estimation process; in the paper: B and A are the goodness-of-fit statistics (see Eqs. (18) and (19)), T_{1n} is the modified Fan's T_1 test statistic (see Eq. (20)).

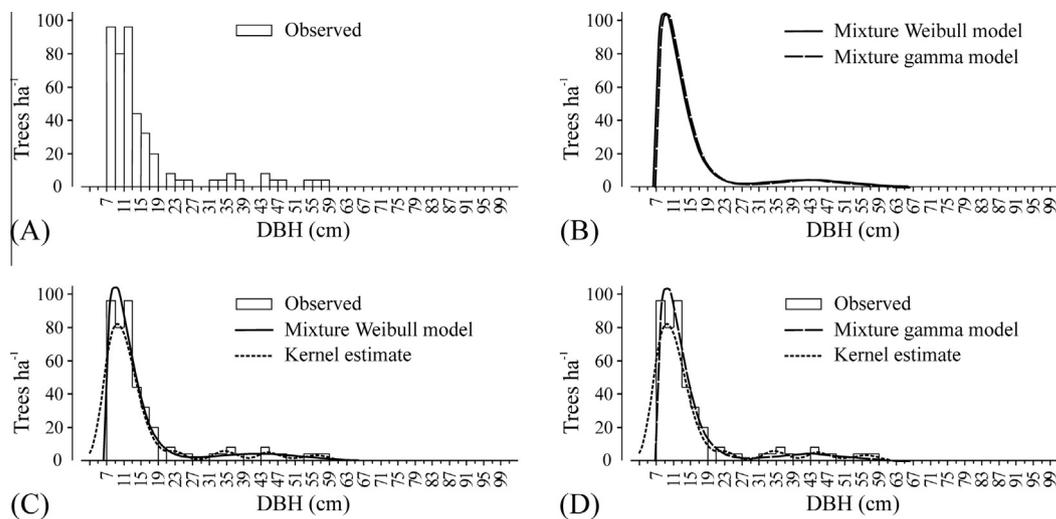


Fig. 4. Approximation of the empirical DBH data (A) for the plot ŠPN30 using the mixture models (B)–(D) and the kernel density estimator (C) and (D). The A statistic (see Eq. (19)) was higher for the kernel density estimator than for the mixture models.

The development of two-cohort, two-storied, multi-species stands from the southern Appalachians was influenced by many

natural and anthropogenic factors. Canopy dynamics were driven by stand initiating disturbances such as large fires and blowdowns

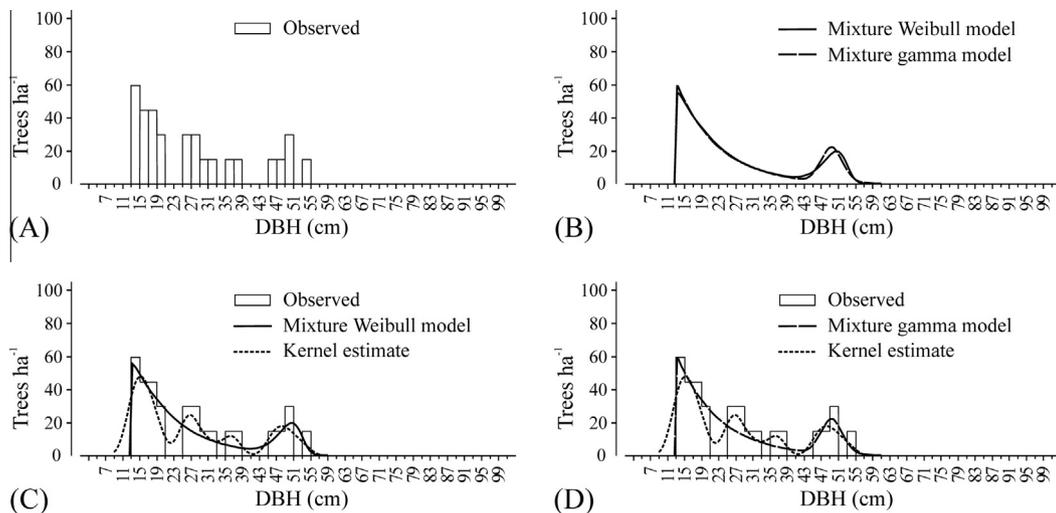


Fig. 5. Approximation of the empirical DBH data (A) for the plot 199016 using the mixture models (B)–(D) and the kernel density estimator (C) and (D). The values of the A statistic (see Eq. (19)) were almost equal for all models.

and small, single or multiple-tree mortality events caused by wind, drought, ice, snow, and lightning strikes (e.g., [65,66,39,49]). Especially, fire was a significant disturbance agent (e.g., [1,2]). Pine and oak species were maintained in presettlement southern Appalachian forests by periodic fire disturbance (e.g., [21,57]). Insects and disease also played an important role (e.g., [64]).

Gap size is a major factor in gap colonisation (e.g., [16]). Shade-intolerant species usually regenerate in large gaps, while a greater proportion of shade-tolerant species will regenerate in small gaps. Patches that were originally created in the gaps can have either loose or dense upper canopy layers at more advanced stages of maturity. Gaps of different sizes were created in southern Appalachian forests, resulting in upper canopy layers of different densities. Light transmission through forest canopies depends primarily on tree density, species and the amount of assimilative apparatus. The resultant two-storied stands can have shade-tolerant dominant species in the upper canopy layer (e.g., red maple) and shade-intolerant dominant species in the lower layer (e.g., sweet birch).

The fir and beech forests of the Świętokrzyskie Mountains were shaped mostly by fine-scale disturbances, and periodically by intermediate-scale disturbances [44]. Fine-scale disturbances were caused mainly by wind, ice, and snow, while intermediate-scale disturbances were connected mostly with the activity of noxious insects attacking trees that were already weakened due to various factors [43]. Contrary to southern Appalachians, fire did not constitute a significant disturbance factor in the Świętokrzyskie Mountains. Fir and beech occur in mixed stands owing to their similar habitat requirements. Both species are considered to be shade-tolerant, although fir is more tolerant to light deficiency than beech. In the Świętokrzyski National Park, in the 20th century, fine-scale and intermediate-scale disturbances led to the creation of gaps in the stand and caused a loosening of canopy closure [44]. The propensity for fir and beech to quickly regenerate both in gaps and in understory environments has resulted in two-cohort stands in which fir and beech occur together in both upper and lower canopy layers [45].

A hypothesis that the DBH components, representing age cohorts and stand layers, can be associated with the dominant species was confirmed for the majority of the stands investigated from the southern Appalachians but was rejected for most stands from the Świętokrzyskie Mountains. The results obtained suggest

that: (1) in multi-species stands from the southern Appalachians, different tree species can be dominant species in specific age cohorts (i.e., they dominate particular age cohorts); (2) in two-species stands from the Świętokrzyski National Park, a tree species usually has a similar proportion in particular age cohorts. These differences are probably caused by the greater number of species and the larger average area of gaps and patches under a loose stand canopy in the southern Appalachians, as opposed to the Świętokrzyskie Mountains [65,67,59,44]. In a larger gap, seedlings and saplings strongly compete and usually one species becomes dominant in a given age cohort. If a stand is composed of a few or even a dozen or so species, different species often dominate the subsequent age cohort. A different mechanism occurs when there are many small gaps and patches under a loose stand canopy, situated close to each other in a forest with two dominant species. In those cases, one of these two species wins in individual gaps and patches under a loose stand canopy and the process repeats in each age cohort. As the result, age cohorts are created composed of the two species occurring in similar proportions.

Parameter estimation can be especially problematic with mixed distributions. Of the various methods for estimating these parameters, the MLE methods have been the most frequently used. These methods are generally regarded as the best because they utilise all information measured for the distribution (e.g., [4,12,56]). The EM algorithm especially with the Newton-type methods is suitable for various samples [9,36]. The method of moments (MM) and least squares (LS) estimation have been also employed. When the variables are random, the MLE methods provide a consistent approach to parameter estimation problems and have desirable mathematical and optimality properties. They become minimum variance unbiased estimators as the sample size increases and approximate sample variances that can be used to generate confidence bounds and hypothesis tests for the parameters [35]. On the other hand the numerical estimations are usually non-trivial. Maximum likelihood estimates can be biased for small samples. In the case of some single and double theoretical distributions the MLE methods are not recommended for small data sets (see e.g., [56,70]). The MLE is usually sensitive to the choice of initial values.

For mixture Weibull model and mixture gamma model, without loss of generality, the parameters can be taken to be the shape and the scale parameter or the mean and SD (e.g., [31]). Knowing the

shape and the scale parameter, one can calculate the mean and SD (and vice versa) for these models. One can generate initial values from a set of empirical data for indirect estimation; (see e.g., [30]). Applying this procedure, we adopt a set of empirical data as a parameter space for the mean. Four simple methods for choosing initial values were presented and analysed. The multistart method is very effective (e.g., [14,36]). There are many variants of this method; the general idea involves choosing a complete set of initial values that cover the data space in a different ways. In the study presented here, a complete set of initial values consisted of 45 subsets for the mean. The min. k /max. k method is an extended version of the classical min/max method (e.g., [53,54]). In the extended version, we introduced additional pairs of initial values for the mean, being the k th smallest and the k th largest values in a data set. We used the 5th and 10th smallest and biggest values in a data set, respectively. This simple extension significantly increased the usefulness of the method. The classical min/max method appeared to be the least efficient of the six methods for choosing the initial values in Podlaski [46,47]. In the present study, the extended version was second to the multistart method. This comparison of the effectiveness of the classical min/max method and its extended version confirms the well-known thesis that the better the initial values cover the data space, the greater is the probability of finding the global extreme. This study did not include all of the simple methods for choosing the initial values in mixture inference; more are described in Böhning [6] and in McLachlan and Peel [35]. The tested methods for choosing initial values did not require additional information, they use only standard data obtained during the measurement of DBHs in a stand (species and DBH). The application of the presented methods enabled the estimation of parameters using both mixture Weibull model and mixture gamma model, though one should remember that problems with estimation of parameters increase when the DBH components of mixture models overlap (which concerns especially mixture Weibull model) [71,17].

Diameter distribution models have been developed using plots of various areas; from 0.04 ha to 10 ha (e.g., [63,19,3]). The diameter distribution shape depends on spatial scale (e.g., [22,25,3]). The number of DBHs measured within a plot is important for reliable estimation of the distribution function [34,70]. Considerable reduction in variance and bias has been found in the Weibull parameters (single model) when sample size changed from 30 to 50, and the further reduction thereafter had a decreasing rate [56]. Analysing the effectiveness of the tested methods for choosing initial values, the following strategy is proposed for maximising the likelihood:

1. For small data sets (number of DBHs measured < 80) the multistart method is proposed – for the two-component mixture Weibull or gamma model select initial values using 10 grid points on the data space (a complete set of initial values consists of 45 subsets should be used).
2. For large samples (number of DBHs measured \geq 80) the multistage method is proposed – for the two-component mixture Weibull or gamma model select initial values using the min. k /max. k (for $k = 1, 5, 10$) and 0.5/1.5/mean methods, run the numerical procedure for each method, and when no two solutions are the same, apply the multistart method also.

The proposed procedure enables, in most cases, the estimation of the parameters of two-component mixtures of either the Weibull or gamma distribution for different number of DBHs measured. In the case of distinct random, local multimodality, one may increase the number of grid points and subsets for initial values for the multistart method and/or number k for the min. k /max. k method.

The presented results show high accuracy of the Weibull and the gamma mixture distributions. This accuracy is similar to or even higher than the precision of approximates obtained using the kernel density estimators. It is a very interesting result because the kernel density estimators are characterised by high flexible properties [27]. Perhaps, these results are an effect of the accurate selection of the parametric models (two-component mixture distributions) to the empirical DBH data. In two-cohort stands the empirical DBH distributions usually reach the first local maximum, then they decrease and increase again, reaching the second local maximum. Two-component mixture models approximate the presented empirical distributions with high accuracy; they precisely fitted fragments around the main “sharp” maxima. Smaller local maxima do not have a significant influence on the quality of the approximation. Kernel density estimators are less precise in smoothing the “sharp” extremes. Non-parametric methods may appear much more useful for approximating the empirical DBH distributions in the multilayered stands.

5. Conclusions

- In mixed-species, two-cohort, two-storied and partly multilayered stands, two-component mixtures of Weibull or gamma functions will model empirical DBHs very well because:
 1. they allow one to determine precisely two local maxima, even if they are not clearly marked, despite the occurrence of random, local multimodality; and
 2. they enable us to determine two subpopulations, connected e.g., with dominant species or age cohorts.
- One of the main limitations in applying two-component models is with problems encountered in parameter estimation. The likelihood function is often irregular and multimodal and that is why the numerical algorithms are unstable. Finally, the global maximum may not be found, or lack of convergence of the estimation process may occur. The closer the initial values adopted in numerical procedures are to the values for which the log likelihood function reaches the global maximum, the more likely it is that we will assess the parameters correctly. Hence, the choice of initial values is of critical importance. Depending on the number of DBHs measured, three methods for choosing the initial values are recommended: min. k /max. k , 0.5/1.5/mean, and multistart.
- In two-cohort, two-storied, multi-species stands with various proportions of dominant tree species, the two-component DBH mixture model can be used to explain relationships between the two-component DBH structure, age cohorts and species composition of stands. The method presented here allows one to carry out this kind of analysis based solely on measurements of DBH. This approach may have additional applications in ecology (e.g., in forest dynamics and forest modelling).

References

- [1] M.D. Abrams, D.A. Orwig, T.E. Demeo, Dendroecological analysis of successional dynamics for a presettlement-origin white-pine mixed-oak forest in the southern Appalachians, *J. Ecol.* 84 (1996) 328.
- [2] M.A. Albrecht, B.C. McCarthy, Effects of prescribed fire and thinning on tree recruitment patterns in central hardwood forests, *For. Ecol. Manage.* 226 (2006) 88.
- [3] A. Alessandrini, F. Biondi, A. Di Filippo, E. Ziaco, G. Piovesan, Tree size distribution at increasing spatial scales converges to the rotated sigmoid curve in two old-growth beech stands of the Italian Apennines, *For. Ecol. Manage.* 262 (2011) 1950.
- [4] R.L. Bailey, T.R. Dell, Quantifying diameter distributions with the Weibull function, *For. Sci.* 19 (1973) 97.
- [5] D.B. Botkin, J.K. Janak, J.R. Wallis, Rationale, limitations, and assumptions of a northeastern forest simulator, *IBM J. Res. Dev.* 16 (1972) 106.

- [6] D. Böhning, Computer-assisted Analysis of Mixtures and Applications Meta-analysis, Disease Mapping and Others, Chapman & Hall/CRC, Boca Raton, 2000.
- [7] T. Bučar, M. Nagode, M. Fajdiga, Reliability approximation using finite Weibull mixture distributions, *Reliab. Eng. Syst. Safety* 84 (2004) 241.
- [8] J.S. Denslow, A.M. Ellison, R.E. Sanford, Treefall gap size effects on above- and below-ground processes in a tropical wet forest, *J. Ecol.* 86 (1998) 597.
- [9] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. R. Stat. Soc. B* 39 (1977) 1.
- [10] J. Du, Combined algorithms for constrained estimation of finite mixture distributions with grouped data and conditional data, Dissertation, McMaster University, 2002.
- [11] J. Emborg, M. Christensen, J. Heilmann-Clausen, The structural dynamics of Suserup Skov, a near-natural temperate deciduous forest in Denmark, *For. Ecol. Manage.* 126 (2000) 173.
- [12] B.S. Everitt, D.J. Hand, *Finite Mixture Distributions*, Chapman & Hall/CRC, London, 1981.
- [13] Y.Q. Fan, Testing the goodness of fit of a parametric density function by kernel method, *Econ. Theory* 10 (1994) 316.
- [14] S. Finch, N. Mendell, H. Thode, Probabilistic measures of adequacy of a numerical search for a global maximum, *J. Am. Stat. Assoc.* 84 (1989) 1020.
- [15] Food and Agriculture Organization, International Soil Reference and Information Centre, and International Soil Science Society, World reference base for soil resources. World Soil Resources Rep No 103. Food and Agriculture Organization, 2006.
- [16] P.C. Goebel, D.M. Hix, Development of mixed-oak forests in southeastern Ohio: a comparison of second-growth and old-growth forests, *For. Ecol. Manage.* 84 (1996) 1.
- [17] J.H. Gove, M.J. Ducey, W.B. Leak, L. Zhang, Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr type III distribution, *Forestry* 81 (2008) 161.
- [18] G. Gratzner, C.D. Canham, U. Dieckmann, A. Fischer, Y. Iwasa, R. Law, M.J. Lexer, T. Spies, B. Splechtna, J. Szwagrzyk, Spatio-temporal development of forests – current trends in field methods and models, *Oikos* 107 (2004) 3.
- [19] C.A. Gronewald, A.W. D'Amato, B.J. Palik, The influence of cutting cycle and stocking level on the structure and composition of managed old-growth northern hardwoods, *For. Ecol. Manage.* 259 (2010) 1151.
- [20] W.L. Hafley, H.T. Schreuder, Statistical distributions for fitting diameter and height data in even-aged stands, *Can. J. For. Res.* 7 (1977) 481.
- [21] R.M. Hubbard, J.M. Vose, B.D. Clinton, K.J. Elliott, J.D. Knoepp, Stand restoration burning in oak-pine forests in the southern Appalachians: effects on aboveground biomass and carbon and nitrogen cycling, *For. Ecol. Manage.* 190 (2004) 311.
- [22] M.K. Janowiak, L.M. Nagel, C.R. Webster, Spatial scale and stand structure in northern hardwood forests: implications for quantifying diameter distributions, *For. Sci.* 54 (2008) 497.
- [23] A. Jaworski, R. Podlaski, Structure and dynamics of selected stands of primeval character in the Pieniny National Park, *Dendrobiology* 58 (2007) 25.
- [24] A. Jaworski, R. Podlaski, Modelling irregular and multimodal tree diameter distributions by finite mixture models: an approach to stand structure characterisation, *J. For. Res.* 17 (2012) 79.
- [25] K. Král, T. Vrška, L. Hort, D. Adam, P. Šamonil, Developmental phases in a temperate natural spruce-fir-beech forest: determination by a supervised classification method, *Eur. J. For. Res.* 129 (2010) 339.
- [26] K.P. Lertzman, B. Dorner, J. Fall, Three kinds of heterogeneity in fire regimes: at the crossroads of fire history and landscape ecology, *Northwest Sci.* 72 (1998) 4.
- [27] Q. Li, J.S. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, 2007.
- [28] Y. Li, R.S. Singh, Y. Sun, Goodness-of-fit tests of a parametric density functions: Monte Carlo simulation studies, *J. Stat. Res.* 39 (2005) 103.
- [29] C. Liu, L. Zhang, C.J. Davis, D.S. Solomon, J.H. Gove, A finite mixture model for characterizing the diameter distribution of mixed species forest stands, *For. Sci.* 48 (2002) 653.
- [30] P.D.M. Macdonald, with contributions from J. Du, *mixdist: Mixture distribution models*, R package version 0.5-1. www.math.mcmaster.ca/peter/mix/mix.html. Accessed 1 April 2010, 2004.
- [31] P.D.M. Macdonald, T.J. Pitcher, Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures, *J. Fish. Res. Board Can.* 36 (1979) 987.
- [32] M. Maltamo, J. Puumalainen, R. Paivinen, 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 (1995) 284.
- [33] J.M. Matuszkiewicz, *Zespoły leśne Polski*, PWN, Warszawa, 2008.
- [34] L. Mehtätalo, C. Comas, T. Pukkala, M. Palahi, Combining a predicted diameter distribution with an estimate based on a small sample of diameters, *Can. J. For. Res.* 41 (2011) 750.
- [35] G.J. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
- [36] G.J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, Wiley, Hoboken, 2008.
- [37] T.A. Nagel, J. Diaci, Intermediate wind disturbance in an old-growth beech-fir forest in southeastern Slovenia, *Can. J. For. Res.* 36 (2006) 629.
- [38] NPS, 2012 National Park Service, USDI. <http://www.nps.gov/grsm/naturescience/index.htm>. Accessed 10 March 2012.
- [39] J.M. Olano, M.W. Palmer, Stand dynamics of an Appalachian old-growth forest during a severe drought episode, *For. Ecol. Manage.* 174 (2003) 139.
- [40] A. Pagan, A. Ullah, *Nonparametric Econometrics*, Cambridge University Press, New York, 1999.
- [41] E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Stat.* 33 (1962) 1065.
- [42] G. Piovesan, A. Di Filippo, A. Alessandrini, F. Biondi, B. Schirone, Structure, dynamics and dendroecology of an Apennine old-growth beech forest, *J. Veg. Sci.* 16 (2005) 13.
- [43] R. Podlaski, Accuracy assessment of a small-area method for estimating the spatial distribution of the degree of tree damage, *Environ. Monit. Assess.* 135 (2007) 339.
- [44] R. Podlaski, Dynamics in central European near-natural *Abies-Fagus* forests: Does the mosaic-cycle approach provide an appropriate model?, *J. Veg. Sci.* 19 (2008) 173.
- [45] R. Podlaski, Diversity of patch structure in central European forests: Are tree diameter distributions in near-natural multilayered *Abies-Fagus* stands heterogeneous?, *Ecol. Res.* 25 (2010) 599.
- [46] R. Podlaski, Modelowanie rozkładów pierśnic drzew z wykorzystaniem rozkładów mieszanych. I. Definicja, charakterystyka i estymacja parametrów rozkładów mieszanych, *Sylwan* 155 (4) (2011) 244.
- [47] R. Podlaski, Modelowanie rozkładów pierśnic drzew z wykorzystaniem rozkładów mieszanych II. Aproksymacja rozkładów pierśnic w lasach wielopiętrowych, *Sylwan* 155 (5) (2011) 293.
- [48] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org. Accessed 30 April 2008, 2008.
- [49] J.S. Rentch, T.M. Schuler, G.J. Nowacki, N.R. Beane, W.M. Ford, Canopy gap dynamics of second-growth red spruce-northern hardwood stands in West Virginia, *For. Ecol. Manage.* 260 (2010) 1921.
- [50] M.R. Reynolds, T. Burk, W.H. Huang, Goodness-of-fit tests and model selection procedures for diameter distribution models, *For. Sci.* 34 (1988) 373.
- [51] M. Rosenblatt, Remarks on some nonparametric estimates of a density functions, *Ann. Math. Stat.* 27 (1956) 832.
- [52] J.R. Runkle, Disturbance regimes in temperate forests, in: S.T.A. Pickett, P. White (Eds.), *The Ecology of Natural Disturbance and Patch Dynamics*, Academic Press, Orlando, 1985, pp. 17–33.
- [53] W. Seidel, K. Mosler, M. Alker, A cautionary note on likelihood ratio tests in mixture models, *Ann. Inst. Stat. Math.* 52 (2000) 481.
- [54] W. Seidel, H. Ševčíková, M. Alker, On the power of different versions of the likelihood ratio test for homogeneity in an exponential mixture model. Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 92. Universität der Bundeswehr, Hamburg, 2000b.
- [55] A. Shamilov, S. Asma, Finite mixtures of MaxEnt distributions, *Appl. Math. Comput.* 206 (2008) 530.
- [56] B.D. Shiver, Sample sizes and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions, *For. Sci.* 34 (1988) 809.
- [57] S.A. Signell, M.D. Abrams, J.C. Hovis, S.W. Henry, Impact of multiple fires on stand structure and tree regeneration in central Appalachian oak forests, *For. Ecol. Manage.* 218 (2005) 146.
- [58] T.M. Smith, D.L. Urban, Scale and resolution of forest structural pattern, *Vegetatio* 74 (1988) 143.
- [59] B.E. Splechtna, G. Gratzner, Natural disturbances in Central European forests: approaches and preliminary results from Rothwald, Austria, *For. Snow Landscape Res.* 79 (2005) 57.
- [60] D. Titterton, A. Smith, U. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.
- [61] USDA, Soil survey of Great Smoky Mountains National Park, Tennessee and North Carolina, United States Department of Agriculture, Natural Resources Conservation Service, 2009.
- [62] T.T. Veblen, D. Kulakowski, K.S. Eisenhart, W.L. Baker, Subalpine forest damage from a severe windstorm in northern Colorado, *Can. J. For. Res.* 31 (2001) 2089.
- [63] C. Westphal, N. Tremer, G. von Oheimb, J. Hansen, K. von Gadow, W. Härdtle, Is the reverse J-shaped diameter distribution universally applicable in European virgin beech forests?, *For. Ecol. Manage.* 223 (2006) 75.
- [64] P.S. White, C.V. Cogbill, Spruce-fir forests of eastern North America, in: C. Eagar, M.B. Adams (Eds.), *Ecology and Decline of Red Spruce in the Eastern United States*, Springer-Verlag, New York, 1992, pp. 3–39.
- [65] P.S. White, M.D. Mackenzie, R.T. Busing, Natural disturbance and gap phase dynamics in southern Appalachian spruce-fir forests, *Can. J. For. Res.* 15 (1985) 233.
- [66] P.S. White, R.R. Buckner, J.D. Pittillo, C.V. Cogbill, High-elevation forests: spruce-fir forests, northern hardwood forests, and associated communities, in: W.H. Martin, S.G. Boyce, J.P. Echternacht (Eds.), *Biodiversity of the Southeastern United States: Upland Terrestrial Communities*, Wiley, New York, 1993, pp. 305–338.
- [67] K.D. Woods, Intermediate disturbance in a late-successional hemlock – northern hardwood forest, *J. Ecol.* 92 (2004) 464.
- [68] S.W. Woudenberg, B.L. Conkling, B.M. O'Connell, E.B. LaPoint, J.A. Turner, K.L. Waddell, *The Forest Inventory and Analysis Database: Database Description and Users Manual Version 4.0 for Phase 2*. Gen. Tech. Rep. RMRS-GTR-245. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2010.

- [69] M. Zasada, Możliwość zastosowania rozkładów mieszanych do modelowania rozkładów pierśnic drzew w naturalnych klasach biosocjalnych, *Sylvan* 147 (9) (2003) 27.
- [70] M. Zasada, Evaluation of the double normal distribution for tree diameter distribution modeling, *Silva Fenn.* 47 (2) (2013). Article id 956.
- [71] M. Zasada, C.J. Cieszewski, 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 (2005) 145.
- [72] L. Zhang, C. Liu, Fitting irregular diameter distributions of forest stands by Weibull, modified Weibull, and mixture Weibull models, *J. For. Res.* 11 (2006) 369.
- [73] L. Zhang, J.H. Gove, C. Liu, W.B. Leak, A finite mixture distribution for modeling the diameter distribution of rotated-sigmoid, uneven-aged stands, *Can. J. For. Res.* 31 (2001) 1654.