Eur J Forest Res (2010) 129:1169–1179
DOI 10.1007/s10342-010-0407-y

ORIGINAL PAPER

Modeling diameter distribution of Austrian black pine (*Pinus nigra* Arn.) plantations: a comparison of the Weibull frequency distribution function and percentile-based projection methods

Tatiana V. Stankova · Tzvetan M. Zlatanov

Received: 30 June 2009 / Revised: 9 November 2009 / Accepted: 14 May 2010 / Published online: 6 July 2010

Abstract The main purpose of the present investigation is to examine and compare three methods for diameter distribution modeling in terms of their fitness to predict from stand level variables the diameter distributions of even-aged Austrian black pine (*Pinus nigra* Arn.) plantations in Bulgaria. The percentile-based projection method involving empirical probability density function based on 12 percentiles was the first method tested. A new modified approach based on the first method was proposed as the second alternative. The third method was the 2-parameter Weibull functional model in which parameters were recovered from the first and the second raw moments and the second central moment of the empirical distributions. The Kolmogorov–Smirnov test was applied to compare the experimental distributions with the predicted ones, and estimation of the error indices was employed to evaluate the total absolute deviation of the predicted numbers from the actual ones by diameter class. The two-parameter Weibull function proved superior to the examined alternative percentile-based projection methods and the newly proposed percentile method, without a driver percentile showed improved precision compared to the classical percentile method (with a driver percentile). The parameters of the Weibull frequency distribution function can be easily recovered from the stand quadratic mean diameter. Consequently, this diameter distribution model could be incorporated as a sub-model for stand horizontal structure characterization within the Stand Density Management Diagram modeling framework.

Communicated by D. Mandallaz.

T. V. Stankova (✉) · T. M. Zlatanov
Forest Research Institute of Bulgarian Academy of Sciences, 132 "St. Kliment Ohridski "blvd., 1756 Sofia, Bulgaria
e-mail: tatianastankova@yahoo.com

Keywords Horizontal stand structure · Moments method for parameter recovery · Structural stand density management diagram

Introduction

Growth and yield models which incorporate diameter distribution submodels characterize the horizontal structure of the stands (sensu Porté and Bartelink 2002). The development of a diameter–distribution model along with height-diameter and taper models would convert the average-tree stand level models to distribution stand level models, which not only model the average and total dimensions of the stand but also partly integrate the natural variability among the trees in a stand, defining different size classes for each modeled characteristic of the stand (Porté and Bartelink 2002). The incorporation of the distribution models into the modeling framework of the stand level models is even needed in order to guarantee the compatibility of growth predictions made by the different model types (Pretzsch et al. 2006). Also, the conversion-based harmonization method derived for adjustment of estimated forest stocking measures to international reference standards (Rennolls et al. 2009) suggested that direct conversion of stand-level variables is possible in principle, if the stand-variable list is augmented to include measures which characterize the diameter distribution, and the stand height curve.

Although very important for the forest management planning and for the evaluation of the outcomes from the forest management activities, diameter distributions are not always available in the forest inventories and are not predictable by the Growth and Yield Tables, which are built as average-tree stand models (e.g. in Bulgaria). To overcome this shortage of information, diameter distribution models are being developed to predict the pattern of tree diameter distribution from stand level variables, information of which is usually readily available in the forest inventories. Numerous functional relationships such as beta, gamma, lognormal, normal, Weibull and Johnson's S_B have been examined for their applicability in modeling diameter distribution. Each of the statistical distributions traditionally considered for fitting forest mensuration data have strengths and weaknesses which can result in extremes of goodness of fit from one data set to another (Hafley and Schreuder 1977). Dimitrov (2003) found that the beta function best describes the diameter structure of pure natural stands of Scots pine (*Pinus sylvestris* L.), White fir (*Abies alba* Mill.) and Norway spruce (*Picea abies* (L.) Karst.) in Bulgaria, because of its ability to predict successfully both symmetric and asymmetric distributions. Hafley and Schreuder (1977), on the other hand, compared 6 (beta, normal, Jonson's S_B, Weibull, gamma and lognormal) types of diameter distribution functions in terms of their flexibility and concluded that Johnson's S_B was the best performing model when fitted to a variety of sample distributions. Weibull frequency distribution function has gained the most popularity, because of its relative simplicity and ability to describe a wide range of unimodal distributions and its applicability to monospecific even-aged conifer stands have been revealed (Newton et al. 2005; Álvarez-González et al. 2002; Diéguez-Aranda et al. 2006; Castedo-Dorado et al. 2007).

According to Borders et al. (1987), a common thread tying the diameter distribution models is their predefined functional form. As an alternative to the functional distribution models, they proposed an empirical probability density function based on 12 percentiles, which should be able to represent a range of diameter distributions with different shape and modality.

The main objectives of the present investigation are to (1) examine the percentile method by Borders et al. (1987), a new modification of the percentile method application and the 2-parameter Weibull function in terms of their goodness of fit to predict from stand level variables the diameter distributions of even-aged Austrian black pine (*Pinus nigra* Arn.) plantations, and (2) given (1) derive recommendations for their practical application in further modeling activities.

Materials and methods

Data sets

The data set used to fit the proposed distribution models consisted of 140 diameter frequency distribution measurements obtained by temporary variable–sized sample plots of circular or rectangular form, which were established in Austrian black pine plantations during the 2002–2005 period (Table 1). The plots were situated in the mountainous part of south-western Bulgaria at altitudes from 275 to 1350 m a. s. l. and on slopes ranging from 0° to 39°. The stands within which the plots were established covered the variety of sites, densities and growth stages of Austrian black pine plantations in this area of Bulgaria. For each plot, all breast height diameters (dbh $\geq$ 1 cm) to the nearest centimeter and the top tree heights (with precision of 0.25 m) were measured and subsequently used to calculate the following stand variables: quadratic mean diameter

Table 1 Stand and tree characteristics of the fit and the validation data sets for modeling the diameter distributions of Austrian black pine plantations in Bulgaria

	Variable	Fit data set (N = 140) n = 6764		Validation data set									
				Subset 1 (N = 5) n = 893		Subset 2 (N = 9) n = 390		Subset 3 (N = 12) n = 1748		Subset 4 (N = 15) n = 1482		Total (N = 41) n = 4513	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Stand level	QMD (cm)	17.0	6.8	9.1	1.3	25.6	3.4	25.9	3.1	10.4	3.4	18.1	8.5
	Density (ha^{-1})	2836	1737	4171	1749	872	171	1270	402	7061	1782	3655	3050
	Age (years)	45	16	25	0	93	5	54	11	17	5	46	30
	Dominant height (m)	15.8	4.7	10.4	0.3	18.0	3.5	23.4	2.3	9.4	2.8	14.8	6.4
	Plot size (m^2)	245.2	188.4	489.8	265.1	499.7	79.7	1147.5	6.8	141.3	10.6	557.0	423.1
Tree level	DBH (cm)	16.3	7.4	8.6	2.5	24.7	6.7	24.6	6.2	9.6	4.3	16.5	4.9

QMD, quadratic mean diameter; DBH, diameter at breast height; N, number of plots; n, number of trees (cm), density (stems per hectare) and dominant height (m). The plantation age (years) was determined from the plantation establishment records. The stand dominant height was defined as the mean height of the 20% of the tallest trees in the stand and was determined from the sampled top tree heights in each plot. Validation data was derived from published or granted data sets and consisted of 4 sub-sets (Table 1). Subset 1 includes five temporary sample plots in 25-year-old plantations published by Marinov et al. (1997), while Subset 2 contains 9 temporary sample plots in advanced age (82–97 years) Austrian black pine plantations (Zlatanov and Velichkov, unpubl.). Subset 3 consists of 4 forest inventory plots measured 3 times at 10 year periods, while Subset 4 contains 3 sample plots measured 5 times at different time intervals (Marinov 2006).

Diameter distribution models

Three distribution modeling methods were examined. The first method was the percentile-based method introduced by Borders et al. (1987), which defines the empirical probability density function with 12 ordered percentiles expressed by regression equations on the stand level variables. The percentile which can be predicted with the highest confidence is selected as the driver percentile and a system of equations is formulated around it. The driver percentile is expressed as a function of the stand level variables alone, while the other percentiles are expressed as a function of stand level variables and information pertaining to the adjacent percentile. The 12 percentiles which are evaluated include the following (Borders et al. 1987): 0-th, 5-th, 15-th, 25-th, 35-th, 45-th, 55-th, 65-th, 75-th, 85-th, 95-th and the 100-th.

A modified application of the percentile-based projection method was introduced in the present study as a second method after preliminary evaluation of the individual percentile equations. The preliminary test revealed that the equations are consistent with each other across all diameter classes and without illogical crossovers suggesting that no constraint, through a relationship to a driver percentile, is required. Thus, the modified percentile method estimates all 12 percentiles as functions of the stand level variables alone.

As a result, the first two methods produce estimation of the number of trees in each size class without predefined functional form of the distribution:

$$\left| \begin{array}{l} N_K = \left[\dfrac{P_i - D_K^L}{P_i - P_{i-1}}(t_i - t_{i-1}) + (t_j - t_i) + (t_{j+1} - t_j)\dfrac{D_K^U - P_j}{P_{j+1} - P_j}\right] N \\ P_{i-1} < D_K^L < P_i \\ P_j < D_K^U < P_{j+1} \end{array} \right. \quad (1)$$

where N is the total number of trees; N_K is the number of trees in the K-th diameter class; D_K^L and D_K^U are the lower and the upper limits of the K-th diameter class, respectively; P_m is the m-th percentile ($\{i, i-1, j, j+1\} \in m$) and t_m is the proportion of the trees with diameter less than the specified P_m; $m\{0, 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 100\}$.

The stand variables, dominant height, age, density and quadratic mean diameter, were tested to express the percentiles, which was done for each equation by stepwise multiple regression. Using only the parameters which were significant at level $P < 0.05$, the equations were then re-estimated by simultaneously refitting them collectively by seemingly unrelated regression (SUR).

The third distribution modeling method examined was the 2-parameter Weibull probability density function, the coefficients of which were estimated by applying the moment method of the parameter recovery approach:

$$f(x) = \left(\frac{c}{b}\right)\left(\frac{x}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c} \quad (2)$$

where the scale parameter b and the shape parameter c are recovered from the first raw (m_1) and the second central (m_2) moments of the distribution, as proposed in the studies by Diéguez-Aranda et al. (2006) and Castedo-Dorado et al. (2007), through the relationships:

$$b = \frac{m_1}{\Gamma\left(1 + \frac{1}{c}\right)} \quad (3)$$

$$m_2 = \frac{m_1^2}{\Gamma^2\left(1 + \frac{1}{c}\right)}\left(\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right)\right) \quad (4)$$

where Γ is the Gamma function.

The first raw moment m_1, which is the arithmetic mean diameter, was estimated by linear regression on the quadratic mean diameter (QMD), which is the second raw moment:

$$m_1 = a + bQMD \quad (5)$$

as proposed by Stankova et al. (2002), while the second central moment m_2, which is the variance of the distribution, is estimated by the arithmetic and the quadratic mean diameters (Diéguez-Aranda et al. 2006; Castedo Dorado et al. 2007):

$$m_2 = QMD^2 - m_1^2 \quad (6)$$

Models validation and verification

The goodness of fit for each of the examined methods was first estimated by evaluation of the coefficients of determination (R^2), relative mean squire errors (RMSE), and significance levels of the regression model (F-test) and its coefficients (t-tests).

The Kolmogorov–Smirnov test, the test statistic of which is based on the maximum absolute difference between the observed cumulative distribution functions of two samples and is compared to a critical value that should not be exceeded at given significance level ($P < 0.05$ in the present study), was applied to collate the experimental distributions with the predicted ones by each of the proposed models (Little 1983; Liu et al. 2004). Because the Kolmogorov–Smirnov test is sensitive to differences in location and shape, centering of the distributions around their mean by subtracting their location (the mean) was done prior to the analysis. The error indices were estimated as the sums of the absolute differences between predicted and observed number of trees per 1000 m^2 within each diameter class (Reynolds et al. 1988; Liu et al. 2004). The Kolmogorov–Smirnov test statistics (K–S) and the error indices (EI) were determined for both the fit and the validation data sets, as well as for the data subsets as divided according to the type of the distribution, unimodal or multimodal. The test statistics, K–S and EI, were further compared via the Wilcoxon signed-ranks test in order to find statistically significant differences between the examined models, regarding their goodness of fit. In the Wilcoxon test, which allows testing for differences between paired scores when the assumptions required by the paired-samples *t* test cannot be made, test ranks are assigned to each pair of the compared data sets. They are based on the absolute value of the difference between the two test variables, and the sign of the difference is used to classify cases into one of three groups: negative ranks, positive ranks and ties (equal to 0). The similar values of the sums of the negative and the positive ranks suggest similarity in the medians of the compared samples. The estimated test statistic, *Z*, is a standardized measure of the distance between the rank sum of the negative group and its expected value (half the sum of all ranks) where a higher absolute value suggests a significant deviation between the two compared samples. The two-tailed asymptotic significance estimates the probability of obtaining a *Z* statistic that is as extreme or more extreme in absolute value then the one displayed, if there is truly no difference between the group ranks. Mann–Whitney test was also applied to compare the K–S and EI values estimated for the unimodal and the multimodal distributions by the same distribution model.

Results

Parameterization of the distribution models

Through the application of stepwise multiple regression analysis, most of the 12 ordered percentiles were expressed by linear relationships on the quadratic mean diameter among the stand level variables (Table 2). Stand age was included as an independent variable only in the equation for the 100-th percentile in the case of the method without the driver percentile, while the 25-th, the 65-the and the 95-th percentiles were expressed by both, the age and the quadratic mean diameter, in the method with the driver percentile. The Breusch-Pagan test of independence proved the necessity of applying SUR by producing highly significant test statistics for both methods ($\mathrm{chi}^2 = 1721.495$, $P < 0.001$ for the method without the driver and $\mathrm{chi}^2 = 246.425$, $P < 0.001$ for the method with the driver). The individual regression equations and their coefficients, estimated for the percentiles without the driver percentile, were statistically significant, with coefficients of determination ranging from 0.84 to 0.99 (Table 2).

The 55-th percentile was derived as a driver in the application of the method by Borders et al. (1987). To obtain parameter estimates for the system of 12 equations, the left-hand side of each equation (i.e. the dependent variable *y* in Table 2), except for the driver percentile, was rewritten as the difference between adjacent percentiles. All regression equations fitted to the system with the driver percentile by SUR were statistically significant with R^2s ranging from 0.59 to 0.99.

Strongly determined ($R^2 = 0.99$), statistically significant linear relationship of the arithmetic mean on the quadratic mean diameter was derived for the application of the parameter recovery approach by moments for the Weibull 2-coefficient density frequency function (Table 2).

Precision evaluation and comparison of the distribution models

The Kolmogorov–Smirnov test applied to the fit data set showed that the investigated diameter distributions were expressed successfully by the 2-parameter Weibull function in 98% of the cases at $P < 0.05$ significance level. The two percentile methods represented confidently the observed distributions in 93 and 97% of the cases for the method with and without driver percentile, respectively. The means of the Kolmogorov–Smirnov test statistics had values 0.522, 0.625 and 0.703 and the estimated mean error indices were 126, 128 and 137 for the Weibull function, the percentile model without and the percentile model with driver, respectively (Table 3). Ninety-five of the distributions were classified as unimodal, while only 45 exhibited a multimodality pattern. The unimodal distributions had smaller K–S values, while their error indices were higher for all of the investigated methods.

The results from the Wilcoxon signed-ranks test using the entire fit data set revealed statistically significant differences between all examined methods for both the error

Table 2 Regression estimations for the three diameter distribution modeling methods and goodness of fit statistics

Dependent variable y	Regression model									Significance of the regression model		
	a	SE of a	t value[a]	b	SE of b	t value[a]	c	SE of c	t value[a]	R^2	RMSE	P-value[a]
Percentile method without driver percentile: $y = a + bQMD + c(1/age)$												
P_{100}	5.558	0.857	6.48***	1.190	0.033	35.62***	−31.557	14.656	−2.15*	0.928	2.297	1816.86***
P_{95}	2.443	0.279	8.75***	1.173	0.015	77.37***				0.977	1.199	5986.87***
P_{85}	1.162	0.212	5.48***	1.122	0.012	97.04***				0.984	0.941	9417.62***
P_{75}	0.582	0.155	3.75***	1.081	0.009	127.45***				0.991	0.703	16243.87***
P_{65}				1.052	0.003	308.18***				0.999	0.739	94973.06***
P_{55}	−0.534	0.119	−4.48***	1.031	0.007	157.66***				0.994	0.547	24857.10***
P_{45}	−0.803	0.143	−5.60***	0.993	0.008	127.57***				0.992	0.615	16273.31***
P_{35}	−0.912	0.176	−5.18***	0.945	0.010	98.51***				0.985	0.777	9703.91***
P_{25}	−1.534	0.200	−7.69***	0.924	0.011	85.19***				0.981	0.867	7257.74***
P_{15}	−1.809	0.304	−5.96***	0.861	0.017	52.08***				0.949	1.343	2712.37***
P_{5}	−2.071	0.345	−6.01***	0.767	0.019	40.74***				0.918	1.547	1659.84***
P_{0}	−2.149	0.432	−4.98***	0.671	0.024	28.40***				0.843	1.977	806.56***
Percentile method with driver percentile P_{55}: $y = a + bQMD + c(1/age)$												
P_{100}–P_{95}				0.139	0.009	15.55***				0.637	1.937	241.87***
P_{95}–P_{85}				0.101	0.008	12.63***	13.188	4.933	2.67**	0.765	1.199	448.77***
P_{85}–P_{75}				0.070	0.004	19.86***				0.741	0.761	394.23***
P_{75}–P_{65}				0.059	0.004	15.29***				0.629	0.843	233.75***
P_{65}–P_{55}				0.040	0.004	8.99***	7.355	2.327	3.16**	0.593	0.757	202.64***
P_{55}	−0.236	0.058	−4.07***	1.016	0.004	262.65***				0.994	0.553	68984.48***
P_{45}–P_{55}				−0.052	0.003	−16.09***				0.652	0.696	258.88***
P_{35}–P_{45}				−0.054	0.003	−19.14***				0.726	0.609	366.43***
P_{25}–P_{35}				−0.042	0.004	−9.85***	−8.845	2.534	−3.49***	0.686	0.664	302.89***
P_{15}–P_{25}				−0.077	0.004	−18.49***				0.713	0.900	342.03***
P_{5}–P_{15}				−0.108	0.005	−21.04***				0.762	1.114	442.67***
P_{0}–P_{5}				−0.100	0.007	−15.22***				0.627	1.417	231.79***
Two-coefficient Weibull function estimated through parameter recovery method by moments: $y = a + bQMD$												
Arithmetic mean DBH	−0.345	0.047	−7.28***	0.995	0.003	388.647***				0.999	0.203	151046.45***

[a] Level of significance: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; *SE* standard error, *RMSE* relative mean squire error, P_i *i*-th percentile of the distribution, *QMD* quadratic mean diameter, *DBH* diameter at breast height

Table 3 Error index and Kolmogorov–Smirnov statistic for the fit and the validation data sets

Data set	Distribution models	Unimodal distributions				Multimodal distributions				All distributions			
		Kolmogorov–Smirnov statistic		Error index		Kolmogorov–Smirnov statistic		Error index		Kolmogorov–Smirnov statistic		Error index	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Fit data set	*P*	0.608	0.249	131	74	0.662	0.321	122	79	0.625	0.274	128	75
	Pd	0.683	0.338	140	89	0.745	0.396	130	95	0.703	0.358	137	90
	Wb	0.514	0.192	129	76	0.539	0.302	118	75	0.522	0.232	126	76
Validation data set	*P*	0.683	0.268	171	171	0.679	0.267	146	129	0.681	0.264	158	149
	Pd	0.791	0.240	160	159	0.641	0.197	144	129	0.715	0.229	152	143
	Wb	0.490	0.156	155	153	0.471	0.162	135	115	0.480	0.157	145	134

P, percentile method without the driver percentile; Pd, percentile method with the driver percentile; Wb, Two-coefficient Weibull function estimated through parameter recovery method by moments; SD, standard deviation index and the K–S test statistic (Table 4). The Weibull two parameter distribution model was superior to both percentile methods (K–S: $Z = -4.281$ ($P < 0.001$), $Z = -5.45$ ($P < 0.001$); EI: $Z = -2.148$ ($P < 0.032$), $Z = -4.422$ ($P < 0.001$); Table 3), while the percentile method without the driver exhibited statistically significant (K–S: $Z = -2.020$ ($P < 0.043$); EI: $Z = -2.270$ ($P < 0.023$)) superiority to the method by Borders et al. (1987) (Table 3). The Weibull function proved superior to the other two models for the K–S statistics for both types of distributions, but its error index values were significantly better than the method with the driver percentile only (Tables 3, 4). A comparison by Mann–Whitney test between EI and K–S statistics estimated by each method for the subsets with different modality revealed equal ability ($P > 0.05$) of all methods in fitting unimodal and multimodal distributions. All models showed similar pattern of distribution of their error indices, indicating an analogous manner in which each of them captured the variation in the diameter frequency.

An example for the application of the 3 tested models to a sample plot from the fit data set is illustrated in Table 5 and Fig. 1b. The examined diameter distribution is bimodal, with 2 close-to-each other peaks, and was approximated through unimodal distributions by all three methods (Fig. 1b). The percentile methods produced very similar values for the twelve estimated percentiles (Table 5), which deviated from the predictions by the 2-parameter Weibull function only for the upper percentile range (Fig. 1b). All models described the shape of the examined distribution well (K–S test statistics did not exceeding the critical values at $P < 0.05$) and the total absolute deviations of the predicted from the experimentally determined numbers of trees were below the average EI determined, for the fit data set (Tables 3, 5).

Figure 1 shows three example distributions from the fit data set. The first plot (Fig. 1a) reveals highly irregular diameter structure which all methods failed. The second and the third plots (Fig. 1b, c) display bimodal distributions with very close-to-each other peaks, which were successfully approximated by Weibull 2-parameter function. The percentile methods, on the other hand, though attempting to capture the bimodality in the diameter distribution of the plot on Fig. 1c, failed to approximate satisfactorily its pattern and showed tendency to increase the peakedness (kurtosis) of the diameter distribution (Fig. 1c).

Twenty and twenty-one of the distributions of the validation data set were unimodal and multimodal, respectively. Three of the seven permanent sample plots changed their pattern of diameter distribution from unimodal to multimodal over time and one permanent sample plot went from multimodal to unimodal to multimodal in the process

Table 4 Results of Wilcoxon signed-ranks test for comparison of the Error index values and Kolmogorov–Smirnov statistics for the three diameter distribution models for the fit and the validation data sets

Data set	Compared models	Unimodal distributions				Multimodal distributions				All distributions			
		Kolmogorov–Smirnov statistic		Error index		Kolmogorov–Smirnov statistic		Error index		Kolmogorov–Smirnov statistic		Error index	
		Z statistic	Level of significance	Z statistic	Level of significance	Z statistic	Level of significance	Z statistic	Level of significance	Z statistic	Level of significance	Z statistic	Level of significance
Fit	P vs. Pd	−1.615	0.106	−1.617	0.106	−1.127	0.260	−1.724	0.085	−2.020	0.043	−2.270	0.023
	P vs. Wb	−3.244	0.001	−1.648	0.099	−2.992	0.003	−1.336	0.182	−4.281	0.001	−2.148	0.032
	Pd vs. Wb	−4.255	0.001	−3.403	0.001	−3.497	0.001	−2.833	0.005	−5.455	0.001	−4.422	0.001
Validation	P vs. Pd	−1.978	0.048	−1.489	0.136	−0.471	0.638	−0.044	0.965	−0.889	0.374	−1.253	0.210
	P vs. Wb	−2.691	0.007	−2.055	0.040	−3.352	0.001	−1.611	0.107	−4.373	0.001	−2.562	0.010
	Pd vs. Wb	−3.413	0.001	−0.935	0.350	−2.768	0.006	−1.061	0.289	−4.413	0.001	−1.401	0.161

P, percentile method without driver percentile; Pd, percentile method with driver percentile; Wb, two-coefficient Weibull function estimated through parameter recovery method by moments

Table 5 Example of diameter distribution estimation by the 3 alternative models

Plot data: plot size = 231.82 m^2; QMD = 18.2 cm; age = 48 years; tree number = 54				Experimental and predicted diameter distributions:					
Percentile methods				DBH (cm)	n (experim.)	n (P)	n (Pd)	Freq (Wb)	n (Wb)
t_m	P_m	ΔPd_m	Pd_m	10	2	0	0	0.016	1
1	26.6	2.5	26.1	11	2	1	1	0.024	1
0.95	23.8	2.1	23.6	12	5	2	2	0.034	2
0.85	21.6	1.3	21.5	13	3	2	2	0.045	2
0.75	20.3	1.1	20.2	14	6	3	3	0.058	3
0.65	19.1	0.9	19.1	15	3	4	5	0.071	4
0.55	18.2		18.2	16	2	5	5	0.084	5
0.45	17.3	−0.9	17.3	17	3	5	5	0.094	5
0.35	16.3	−1.0	16.3	18	5	5	5	0.100	5
0.25	15.3	−0.9	15.4	19	4	5	5	0.100	5
0.15	13.9	−1.4	14.0	20	5	4	4	0.094	5
0.05	11.9	−2.0	12.0	21	4	4	4	0.081	4
0	10.1	−1.8	10.2	22	2	2	4	0.065	3
Weibull 2-parameter distribution model				23	2	2	2	0.047	3
				24	1	1	1	0.030	2
$m_1 = 17.8$		$c = 5.198$		25	0	1	1	0.017	1
$m_2 = 15.4$		$b = 19.301$		26	2	1		0.009	0
				27	3			0.004	0

QMD, quadratic mean diameter; DBH, diameter at breast height; t_m, proportion of the trees with diameter less than the m-th percentile (P_m and Pd_m); ΔPd_m, difference between the adjacent percentiles for the method with driver, as fitted by the regressions in Table 2; Pd_m, percentiles estimated by the method with the driver; P_m, percentiles estimated by the method without the driver; n (experim.), n (P), n (Pd) and n (Wb), numbers of trees: experimental and as estimated by the 3 different methods; Freq (Wb), diameter frequency estimated by Weibull 2-parameter distribution model

of plot re-measuring. The superiority of the Weibull 2-coefficient function was manifested for the validation data set as well, where the mean K–S had values of 0.480, 0.681 and 0.715, and the mean EI values of 145, 158 and 152, for the Weibull function, the percentile model without and the percentile model with the driver percentile, respectively (Table 3). The K–S values estimated for the Weibull 2-parameter function were significantly lower than those obtained for the percentile methods regardless of the degree of modality, however, no significant difference between the K–S values by the two percentile methods was distinguished (Table 4). The cumulative amounts of the deviations from the observed distributions, as estimated by the EI, however, did not differ significantly between the examined models in most of the cases. The only exception was the percentile method without the driver which had significantly higher values of EI compared to Weibull function (Table 4). Except for one diameter distribution which failed to be represented by the percentile method without the driver, all other distributions from the validation data set were successfully described by all three models.

Discussion

Sixty-eight percent of the diameter distributions within the examined Austrian black pine plantations in Bulgaria were unimodal, thus confirming the general finding that even-aged stands tend to have unimodal diameter distributions that exhibit some degree of skewness (Bailey and Dell 1973; Bullock and Burkhart 2005). The predominating unimodality in the fit data set would influence inevitably the estimation of the equations characterizing the percentiles by both of the percentile methods. The prediction patterns of all distribution models showed similarity to each other which was also indicated by similarity in their error index distributions (Figs. 1, 2).

Analogous results were obtained for the goodness of fit analysis of the models regardless of the degree of modality, demonstrating the superiority of the 2-parameter Weibull model. For most of the evaluated distributions, the percentile-based methods predicted narrower diameter distribution ranges with 0-th and 100-th percentiles higher and lower than the actual ones, respectively (Fig. 1c). Two types of multimodality were distinguished in the data set:

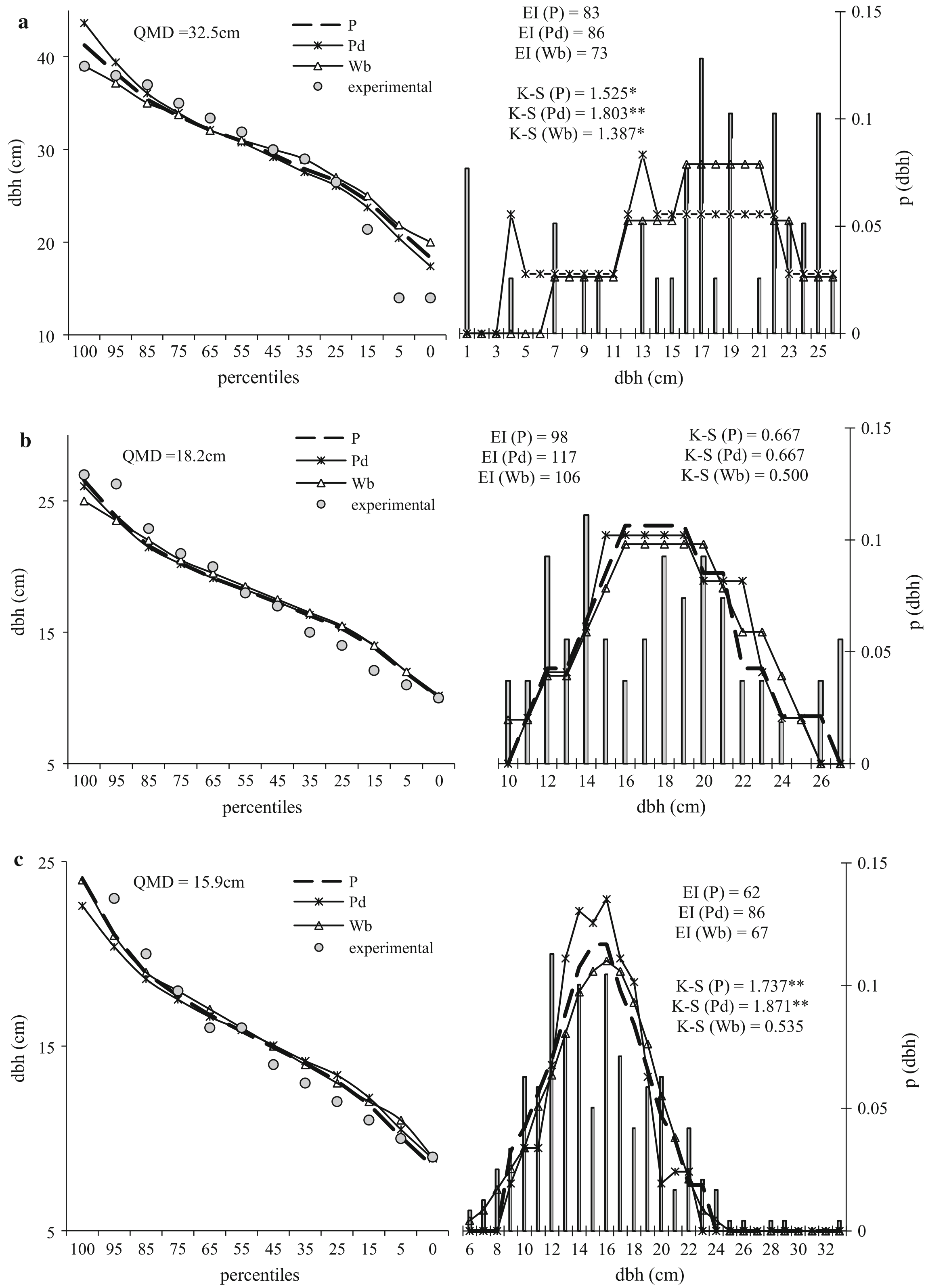

Fig. 1 Estimated values of the 12 main percentiles and diameter distribution predictions for 3 example plots (**a**, **b**, **c**) from the fit data set. Abbreviations: QMD, quadratic mean diameter; P, percentile method without the driver percentile; Pd, percentile method with the driver percentile; Wb, 2-parameter Weibull model; EI (), error index and K–S (), Kolmogorov–Smirnov test statistics by models. Level of significance of the deviation of the modeled distribution from the experimental one according to K–S test: ** $P < 0.05$, *$P < 0.01$

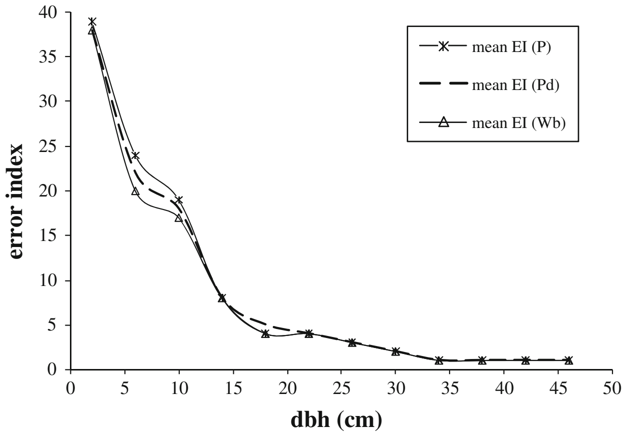

Fig. 2 Error indices (EI) averaged by diameter classes (by 4 cm) by distribution prediction method: percentile method without the driver percentile (P), Percentile method with the driver percentile (Pd) and 2-parameter Weibull model (Wb)

multimodality due to extreme irregularity of the diameter distribution (Fig. 1a) or narrow-ranged distributions with two close to each other peaks (Figs. 1b, c). The second type of multimodality is approximated by the unimodal distribution models. Irregularity, on the other hand, possesses a high degree of randomness, which can be difficult to capture not only by a model of predetermined functional form, but also by any deterministic model such as the percentile-based projection methods. Despite this fact, all tested models exhibited a high degree of precision: the 2-parameter Weibull function, the percentile method without the driver, and the percentile method with the driver, being imprecise in only 3, 5 and 10 out of 181 cases, respectively.

Different methodological approaches have been proposed to estimate the parameters of the Weibull distribution function from stand level variables. Among them, the parameter recovery approaches have been proven more reliable than the parameter estimation ones (Liu et al. 2004), given that the parameter estimation approach relies on the direct relationships of the parameters to selected stand characteristics which can be imprecise and biased (Merganič and Sterba, 2006). Alternatively, the parameter recovery approach by moments which obtains the parameters of the Weibull function from the moments of the distribution has proved to be successful for both the 3-parameter (Liu et al. 2004) and 2-parameter Weibull functions (Diéguez-Aranda et al. 2006; Castedo-Dorado et al. 2007; Merganič and Sterba, 2006). Such an approach was used by Diéguez-Aranda et al. (2006) and Castedo-Dorado et al. (2007) to develop diameter-distribution sub-model of a disaggregation system for inclusion in a stand growth models for Scots pine (*Pinus sylvetris* L.) and Radiata pine (*Pinus radiata* D. Don.) plantations in North-Western Spain. The first raw moment, m_1, in their studies was modeled as the difference between the quadratic mean diameter and exponential function of the stand level variables (density, dominant height, site index and age), and the parameters of the 2-parameter Weibull functions were subsequently estimated by iterative procedures (Eq. 4) and substitution (Eq. 3). Merganič and Sterba (2006), on the other hand, proposed the parameter recovery approach based on relationships of the Weibull function parameters to the mean diameter and the coefficient of variation which proved it applicable to both natural (Norway spruce dominated) and managed (mixed Norway spruce–Scots pine) stands. A complete representation of the Weibull function parameters by the first and the second moments of the distribution was also implemented in the present study (Eqs. 3, 4) involving relationship between the first and the second raw moments (Eq. 5) and their relation to the second central moment (Eq. 6). This approach is based on consistent mathematical relationships and ensures unbiased parameter estimates.

Borders et al. (1987) considered the management history of the studied stands by including in the percentile equations a dummy variable to distinguish between thinned and unthinned stands. As stated by Álvarez-González et al. (2002), thinnings have the greatest impact on the stand structure by affecting diameter growth rates. The expected outcome of a thinning activity would be the narrowing of the diameter range, as found by Borders et al. (1987). An expected outcome from thinnings from below or from above would be also a change in the skewness of a distribution, while a combined type of thinning will probably cause alteration in the distribution kurtosis. Neither of the thinning methods, however, is expected to change unimodal to multimodal distributions and even the opposite effect can be expected. The observed irregularity of approximately 20% of the fit data plots in this study cannot be attributed to human interference, given the known history of the forests. Thus, although the stand management history was not included as a factor in the models of the present investigation, this was not an insuperable obstacle to parameterize the models with relatively high precision.

Borders and Patterson (1990) also excluded the management activity from their percentile prediction equations in a study comparing three methods for diameter distribution modeling. Their results demonstrated the superiority of the percentile method to the Weibull distribution function recovered through percentiles and hence is at odds to the results reported in the present study. The disagreement is even more pronounced due to the fact that the classical percentile method was inferior to all examined methods in almost all instances. The nomination of the best fitted percentile as a driver and the consequent model constraining through its inclusion in the percentile equations inevitably lowers the degree of the variation explained by the regression models, as demonstrated in this study (e.g., compare the coefficients of determination in Table 2). It should be acknowledged, however, that the goodness of fit of any given model is highly influenced by the particular data set to which it is applied and each model can result in extremes of goodness of fit from one data set to another (Hafley and Schreuder 1977) and hence general conclusions cannot be derived. The advantages of both the Weibull distribution model and the empirical percentile-based method should be considered in any particular case study. Despite the superiority of the 2-parameter Weibull function, the percentile method without the driver can still be preferred, particularly if applied to parameterize multimodal distributions alone. Such non-unimodal distribution patterns could be related to irregular tree distribution arising from heterogeneous microsite conditions (e.g. stony soil), dense plantations which have undergone intensive self-thinning, and/or wind and snow damage.

In this study, the diameters were grouped in diameter classes by 4-cm increments and the error indices for the entire validation data set were averaged by diameter classes (Fig. 2). The general tendency of least precise frequency estimation for the lowest diameter classes (Liu et al. 2004; Merganič and Sterba 2006) was confirmed by the present investigation as well. Despite this fact, Weibull 2-parameter model still showed better goodness of fit for the diameters up to 10 cm than the other two methods (Fig. 2). This agrees with the finding by Bullock and Burkhart (2005) that diameter distributions in juvenile pine stands can be successfully characterized with the two parameter Weibull function, though at a lower degree than the mature ones.

The utility of the stand density management diagrams (SDMD), which are stand level models used primarily to derive density control schedules by management objective, has been largely limited to evaluating management outcomes in terms of mean tree size and stand-level volumetric yields (Newton et al. 2004, 2005). However, the advent of the structural SDMD which incorporates a diameter distribution sub-model has been useful in solving a broad array of forest management problems (e.g., designing thinning prescriptions, evaluating wildlife habitat potential, estimating carbon storage potential, assessing the attainment of biologic diversity objectives (Newton et al. 2004, 2005; Newton 2009)). The present investigation on the diameter structure of Austrian black pine plantations in Bulgaria was preceded by elaboration of a dynamic SDMD (Stankova and Shibuya 2007), which was further verified and proven to be better than the whole stand yield tables (Stankova and Zlatanov 2007). Thus, the incorporation of the diameter distribution model into the framework of the SDMD would significantly improve its modeling applicability by not only characterizing the present, but also by predicting the future horizontal structure. The estimation of the diameter distribution by the Weibull 2-parameter function as proposed here only requires information regarding the quadratic mean diameter thus providing an empirical linkage into the SDMD already developed for Austrian black pine plantations. Although beyond the scope of this study, future research should focus on improving the predictive ability of diameter distribution recovery models by examining alternative parameter estimation procedures, such as, mixed-effects modeling (Gregoire 1987; Castedo-Dorado et al. 2006).

Conclusions

The two-parameter Weibull function proved applicable to predict the diameter distributions of even-aged Austrian black pine plantations in Bulgaria. Its superiority to the examined alternative percentile-based projection methods was confidently revealed regardless of the degree of modality. Its parameters can be easily recovered from the stand quadratic mean diameter, thus providing the foundation for the development of structural stand density management diagram for this species in Bulgaria.

References

Álvarez-González JG, Schröder J, Rodríguez Soalleiro R, Ruíz González AD (2002) Modelling the effect of thinnings on the diameter distribution of even-aged Maritime pine stands. For Eco Manag 165:57–65

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

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

Borders BE, Souter RA, Bailey RL, Ware KD (1987) Percentile-based distributions characterize forest stand tables. For Sci 33:570–576

Bullock BP, Burkhart HE (2005) Juvenile diameter distributions of lonlolly pine characterized by the two-parameter Weibull function. New For 29:233–244. doi:10.1007/s11056-005-5651-5

Castedo-Dorado F, Diéguez-Aranda U, Barrio Anta M, Sánchez Rodríguez M, Von Gadow K (2006) A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. For Eco Manag 229:202–213. doi:10.1016/j.foreco.2006.04.028

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. doi:10.1051/forest:2007023

Diéguez-Aranda U, Castedo-Dorado F, Álvarez-González JG, Rodríguez-Soalleiro R (2006) Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). Ecol Model 191:225–242. doi:10.1016/j.ecolmodel.2005.04.026

Dimitrov E (2003) Modeling of structure, volume and assortments of moderate-age and sub-mature dendrocoenoses of Scots Pine, Norway Spruce and Silver Fir. Simolini 94, Sofia (in Bulgarian)

Gregoire TG (1987) Generalized error structure for forestry yield models. For Sci 33:423–444
Hafley WL, Schreuder HT (1977) Statistical distributions for fitting diameter and height data in even-aged stands. Can J For Res 7:481–487
Little SN (1983) Weibull diameter distributions for mixed stands of western conifers. Can J For Res 13:85–88
Liu C, Zhang SY, Lei Y, Newton PF, Zhang L (2004) Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. Can J For Res 34:2424–2432. doi:10.1139/X04-117
Marinov IT (2006) Investigations on the surface water runoff and eroded soil from different areas in Ihtimanska Sredna gora Mountain. For Sci Sofia 2:77–85 (in Bulgarian)
Marinov IT, Mandev A, Grozeva M, Timousheva S, Markov I, Genov K (1997) Investigation on the hydrological effectiveness against erosion of the coniferous plantations in western Bulgaria. Report on the project with National forestry board. Forest Research Institute of BAS, Sofia (in Bulgarian)
Merganič J, Sterba H (2006) Characterization of diameter distribution using the Weibull function: method of moments. Eur J For Res 125:427–439. doi:10.1007/s10342-006-0138-2
Newton PF (2009) Development of an integrated decision-support model for density management within jack pine stand-types. Ecol Model 220:3301–3324
Newton PF, Lei Y, Zhang SY (2004) A parameter recovery model for estimating black spruce diameter distributions within the context of a stand density management diagram. For Chron 80(3): 349–358
Newton PF, Lei Y, Zhang SY (2005) Stand-level diameter distribution yield model for black spruce plantations. For Eco Manag 209:181–192. doi:10.1016/j.foreco.2005.01.020
Porté A, Bartelink HH (2002) Modelling mixed forest growth: a review of models for forest management. Ecol Model 150: 141–188
Pretzsch H, Biber P, Dursky J, Von Gadow K, Hasenauer H, Kandler G, Kenk G, Kublin E, Nagel J, Pukkala T, Skovgaard JP, Sodtke R, Sterba H (2006) Standardizing and categorizing tree growth models. In: Hasenauer H (ed) Sustainable forest management. Growth models in Europe. Springer, Berlin, Heidelberg, New York, pp 39–57
Rennolls K, Päivinen R, San-Miguel-Ayanz J, Tomé M, Skovsgaard JP, Palahi M, Matthews R (2009) Harmonization of European forest growing stocking data using a model-based conversion approach. For Biom Model Inform Sci 1:1–34
Reynolds MR, Burk TE, Huang W-C (1988) Goodness-of fit tests and model selection procedures for diameter distribution models. For Sci 34(2):373–399
Stankova T, Shibuya M (2007) Stand density control diagrams for Scots pine and Austrian black pine plantations in Bulgaria. New For 34(2):123–141. doi: 10.1007/s11056-007-9043-x
Stankova T, Zlatanov T (2007) Verification of a model of stand density control diagram for Austrian black pine plantations in Bulgaria. Silva Lusitana 15(2):171–188
Stankova T, Shibuya M, Hagihara A (2002) A method for density control of forest plantations. For Sci Sofia 2:27–38