FORESTRY IDEAS, 2010, vol. 16, No 2 (40)

DIAMETER DISTRIBUTION MODEL FOR SCOTS PINE PLANTATIONS IN BULGARIA

Tatiana Stankova[1,2] and Ulises Diéguez-Aranda[1]

[1]Department of Agroforestry Engineering, Higher Polytechnic School, University of Santiago de Compostela, 27002 Lugo, Spain.
E-mail: tatianastankova@yahoo.com
[2]Forest Research Institute – BAS, 132 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria.

UDC 630.5

Received: 19 May 2010
Accepted: 13 May 2011

Abstract

The main objective of this study is to derive a diameter distribution model for Scots pine (*Pinus sylvestris* L.) plantations in Bulgaria, which predicts with high confidence the allocation pattern of the tree diameters from stand level variables. As the investigated stands showed predominantly unimodal distribution pattern, their diameter distributions were characterized by 2-parameter Weibull function. Six methods for its parameter estimation were examined: two methods for parameter recovery through moments (PRM_L and PRM_S), a method for parameter recovery through the stand basal area (PRM_B), two parameter prediction methods (PPM_NLS and PPM_MLE), and a modified parameter prediction method based on a mixed-effect model (PPM_Mixed). An empirical percentile model, not connected to a predefined functional form of the distribution, was fitted for comparison. The choice of the best performing model involved estimation of ranks based on Kolmogorov-Smirnov test and Error Index values for goodness of fit evaluation, against a fit and a validation data sets. Two of the parameter recovery methods (PRM_L and PRM_S) and one of the parameter prediction methods (PPM_NLS) performed best, PRM_S being the overall outperformer and PRM_L being the simplest for application. The empirical percentile model ranked fourth and was not advantageous for representing diameter distributions of Scots pine plantations. The two best models derived (PRM_L and PRM_S) describe well the diameter allocation pattern of the Scots pine plantations in Bulgaria and can be applied to estimate the diameter distributions from stand level variables in a simple and reliable way.

Key words: distribution moments, empirical percentile model, parameter prediction method, parameter recovery method, *Pinus sylvestris*, Weibull distribution function.

Introduction

The diameter distribution models are of particular importance for evaluation of the results from the forest management activities, but they are not always available in the forest inventories and often are not predictable by the whole-stand growth and yield tables, as in Bulgaria. To overcome the shortage of information, diameter distribution models are being developed to predict the pattern of tree diameter distribution from stand level variables. Numerous functional relationships have been examined, the Weibull frequency distribution function being most commonly preferred in fitting unimodal distributions, because of its flexibility in fitting a variety of shapes and degrees of skewness, and relative simplicity of estimating its parameters. The parameter recovery approach, which relates percentiles or moments of the distribution, used to recover the Weibull parameters, to stand level variables, and the parameter prediction approach, which connects the function parameters directly to the stand parameters, are the two main methodological groups applied for estimation of the coefficients of the Weibull distribution function. Different modifications and techniques, characteristic to both groups as well as combinations of the estimation approaches have been proposed, revealing the advantages and the limitations of the different methods (Bailey and Dell 1973, Cao 2004, Liu et al. 2004, Merganič and Sterba 2006, Siipilehto 2009). The main objective of the present study is to derive a diameter distribution model for Scots pine (*Pinus sylvestris* L.) plantations in Bulgaria, which predicts with high confidence the allocation pattern of the tree diameters from stand level variables.

Material and Methods

All plots used for data collection are situated in the mountains of Bulgaria and cover the variety of sites, densities and growth stages of the Scots pine plantations. The data set used for model parameterization is composed of two subsets (Table 1). The first subset consists of 153 diameter frequency distributions obtained by temporary sample plots established in 2002–2005. The second subset includes 83 diameter distributions from published data sources (Krastanov et al. 1980, Marinov 2002, Marinov et al. 1997) of permanent sample plots measured one or more times by 2 to 6-year intervals. The validation data set consists of 166 diameter distributions and includes data recorded in a provenance test plantation in 2007 and published data from the forest inventory plots and other permanent sample plots (Efremov 2006) measured once or more times by 10-year intervals. Beside the diameter distribution data, the stand level variables density (ha^{-1}), basal area ($m^2.ha^{-1}$), dominant height (m), quadratic mean diameter (cm), age (years), and site index were also used in the analyses (Table 1). The stand dominant height, where absent, was estimated through its relationship to the mean height established by Stankova et al. (2006), and the site indices of all plots were determined by the site quality tables for Scots pine stands by growth mode (Mihov 1986) for the third type of growth mode (slowing down growth).

Most of the diameter distributions in the modeling data set exhibited unimodal pattern; for this reason, the 2-parameter Weibull frequency distribution function was employed to characterize the diameter distribution pattern of the Scots pine plantations sampled:

$$f(x)=\left(\frac{c}{b}\right)\left(\frac{x}{b}\right)^{c-1}e^{-\left(\frac{x}{b}\right)^{c}} \quad \text{(Eq. 1)},$$

where x stands for diameter at breast height and b and c are the scale and the shape parameters, respectively. The model coefficients were estimated in 6 different ways (Table 2). The first two methods apply the parameter recovery approach through the first raw (m_1) and the second central (m_2) moments of the

Table 1. Stand and tree characteristics of the modeling and the validation data sets for modeling the diameter distributions of Scots pine plantations in Bulgaria.

	Variable	Modeling data set (P = 193) N = 236, n = 44942						Validation data set (P = 73) N = 166, n = 23889		
		Subset 1 (P = 153) N = 153, n = 9372			Subset 2 (P = 40) N = 83, n = 35570					
		Mean	min	max	Mean	min	max	Mean	min	max
Stand variables	QMD, cm	15.2	3.6	35.3	14.7	5.7	23.2	19.8	6.5	34.7
	Density, ha^{-1}	3164	498	12200	3298	679	9305	1706	393	8640
	Age, years	35	10	78	34	14	50	43	15	80
	Dominant height, m	15.9	3.6	32.6	14.5	6.3	23.2	18.4	6.4	32.5
	Basal area, $m^2.ha^{-1}$	41.17	5.54	72.27	44.73	15.72	104.41	42.80	10.12	67.84
	Site index	31	22	38	29	22	38	31	22	38
	Plot size, m^2	270	85	1269	1658	148	2960	954	386	1989
Tree variable	dbh, cm	13.7	2	36	12.3	1	47	16.9	2	52

QMD – quadratic mean diameter; dbh – diameter at breast height; P – number of plots; N – number of distributions; n – number of trees.

distribution (Table 2, Eqs. 2 and 3). In the first method (PRM_L) the raw moment m_1, which is the arithmetic mean diameter, was estimated by linear regression on the quadratic mean diameter (Stankova et al. 2002) (Table 2, Eq. 4). In the second method (PRM_S) the arithmetic mean diameter was evaluated by a relationship (Table 2, Eq. 5) on the quadratic mean diameter and other stand level variables (Diéguez-Aranda et al. 2006). The second central moment m_2, which is the variance of the distribution, is estimated in both methods by the arithmetic and the quadratic mean diameters (Table 2, Eq. 6). The third method for model parameterization (PRM_B) was adapted from Cao et al. (1982) and employs the measured value of the stand basal area per hectare to recover the parameters iteratively (Table 2, Eq. 7), while expressing parameter b through c by Eq. 2. In the parameter calculation the fitted basal area was allowed to deviate from the experimental one (Eq.7) by no more than 5%. Two parameter prediction methods (PPM_NLS and PPM_ML) and one modified parameter prediction method (PPM_Mixed) for coef-

Table 2. Main formulae and estimated regressions for the examined methods for diameter distribution modeling.

Method*	Formulae for estimation of function parameters			Estimated final regressions
PRM_L	(Eq. 2) $b=\frac{m_1}{\Gamma\left(1+\frac{1}{c}\right)}$	(Eq. 3) $m_2=\frac{m_1^2}{\Gamma^2\left(1+\frac{1}{c}\right)}\left(\Gamma\left(1+\frac{2}{c}\right)-\Gamma^2\left(1+\frac{1}{c}\right)\right)$	(Eq. 4) $m_1=a+bQMD$	$m_1 = 0.9974\ QMD - 0.3700$ ($R^2 = 0.9993$; RMSE = 0.1622)
PRM_S		(Eq. 6) $m_2 = QMD^2 - m_1^2$	(Eq. 5) $m_1 = QMD - \exp(\mathbf{X\beta})$	$m_1 = QMD - \exp(-0.0079A - 0.0143SI - 0.8488N)$ ($R^2 = 0.8488$; RMSE = 0.4092)
PRM_B		(Eq. 7) $BA=\frac{N\pi}{40000}\sum_{i=1}^{n}dbh_i^{\,2}w\left(\frac{c}{b}\right)\left(\frac{dbh_i}{b}\right)^{c-1}e^{-\left(\frac{dbh_i}{b}\right)^c}$		The coefficients b and c are iteratively calculated
PPM_NLS	(Eq.1) $f(x)=\left(\frac{c}{b}\right)\left(\frac{x}{b}\right)^{c-1}e^{-\left(\frac{x}{b}\right)^c}$		$b=f(\mathbf{X\beta})$ $c=f(\mathbf{X\beta})$	$b = 1.0295QMD - 4.0285/A - 0.0001N$ ($R^2 = 0.9993$; RMSE = 1.00) $c = 0.1189QMD + 0.0580Hd + 0.0471SI + 0.0001N$
PPM_ML				$b = 0.9405QMD - 32.7534/A + 0.1116Hd - 0.0540BA$ ($R^2 = 0.9742$; RMSE = 1.00) $c = 0.0900QMD + 26.6279/A + 0.1145Hd + 0.0105BA$
PPM_Mixed	(Eq.8) $f(x)=\left(\frac{c}{b_f+b_r}\right)\left(\frac{x}{b_f+b_r}\right)^{c-1}e^{-\left(\frac{x}{b_f+b_r}\right)^c}$		$b_r=f(\mathbf{X\beta})$	$b_f = 0.5774$; $c=3.3466$; $b_r = 0.0051QMD + 0.0054Hd - 0.000011N - 0.0014A$ ($R^2 = 0.6706$; RMSE = 0.0764)
Empirical Percentile Model	(Eq.9) $\left\lvert\begin{array}{l} N_K=\left[\frac{P_i-D_K^L}{P_i-P_{i-1}}(t_i-t_{i-1})+(t_j-t_i)+(t_{j+1}-t_j)\frac{D_K^U-P_j}{P_{j+1}-P_j}\right]N_T \\ P_{i-1}<D_K^L<P_i \\ P_j<D_K^U<P_{j+1}\end{array}\right.$		$P_k=f(\mathbf{X\beta})$	$P_0 = 0.6250QMD - 39.1850/A - 0.0540Hd$ $P_5 = 0.8037QMD - 40.0203/A - 0.1118Hd + 0.0001N$ $P_{15} = 0.8407QMD - 26.8200/A - 0.06780Hd + 0.3582$ $P_{25} = 0.8552QMD - 16.1550/A - 0.00003N$ $P_{35} = 0.9055QMD - 8.75380/A - 0.00005N$ $P_{45} = 0.9382QMD - 5.9976/A + 0.0164Hd - 0.00005N$ $P_{55} = 1.0057QMD - 0.00005N$; $P_{75} = 1.0929QMD + 0.4137$ $P_{65} = 1.038QMD + 8.821/A + 0.047Hd + 0.00003N - 0.9062$ $P_{85} = 1.1105QMD + 10.1768/A + 0.0644Hd + 0.00002N$ $P_{95} = 1.1859QMD - 5.6796/A + 2.5322$ $P_{100} = 1.3417QMD - 46.7059/A - 0.2009Hd + 7.8921$ ($R^2 = 0.9822$; RMSE = 0.9938)

* Method abbreviations are as defined in the text; Γ - Gamma function; m_1 - first raw moment; m_2 - second central moment; $\mathbf{X}$ - vector of the stand variables; $\boldsymbol{\beta}$ - vector of parameters; BA – stand basal area (m^2/ha); N – stand density (1/ha); n - number of diameter classes; w - width of the diameter class; dbh_i - diameter class; b_f, b_r - fixed and random parts of the coefficient b in the mixed-effect model, respectively; N_T - total number of trees; N_K - number of trees in the K-th size class; D_K^L and D_K^U - lower and upper limits of the K-th diameter size class, respectively; P_k - k-th percentile; t_k - the proportion of the trees with diameter less than P_k; $k\{0, 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 100\}$; $\{i, i-1, j, j+1\} \in k$; SI - site index; A - stand age (years); QMD - quadratic mean diameter (cm); Hd - dominant height (m); R^2 – coefficient of determination; RMSE – root mean square error.

ficient estimation were also tested. The first two follow three main steps of application: (i) parameter estimation by plots, (ii) parameter prediction through stepwise multiple regressions on the stand level variables, and (iii) simultaneous refitting of the selected regressions for *b* and *c* through seemingly unrelated regression (Table 2). PPM_NLS and PPM_ML differ in the first step of the estimation procedure, because while PPM_NLS applies the ordinary non-linear least squares technique, PPM_ML utilizes the maximum likelihood estimation. The PPM_Mixed method employed a non-linear mixed-effect model, in which the Weibull function parameters were considered mixed, i.e. composed of a fixed part (common for all distributions) and a random part specific for each plot (Table 2, Eq. 8). After estimation of the mixed-model parameters, the random parts of the parameters were predicted through regressions on the stand level variables.

A modification without driver percentile of the empirical, percentile-based model introduced by Borders et al. (1987), which is not connected to a predefined functional form of the distribution, was also fitted for comparison. This method defines an empirical probability density function with 12 ordered percentiles (Table 2, Eq. 9), and the principle percentiles are further predicted from the stand level variables by stepwise multiple regressions followed by a seemingly unrelated regression.

Kolmogorov-Smirnov (K-S) test was applied to compare the experimental distributions with the predicted ones by each of the proposed models (Little 1983, Liu et al. 2004). To allow comparison of the distribution shapes, their mean value was subtracted prior to the analysis (Stankova and Zlatanov 2010). The Error Index (EI) was estimated for each diameter distribution as the sum of the absolute differences between predicted and

Table 3. Mean values of the comparison test statistics (K-S and EI) and method ranking according to the results of the Wilcoxon rank-sign test*.

Method	Modelling data set				Validation data set				Rank sum		
	K-S		EI		K-S		EI		K-S	EI	Total
	Mean value	Rank	Mean value	Rank	Mean value	Rank	Mean value	Rank			
PRM_L	0.2680	4	119	1	0.1636	1	60	1	5	2	7
PRM_S	0.1942	1	123	2	0.1649	1	62	2	2	4	6
PRM_B	0.2484	3	141	4	0.2642	5	76	3	8	7	15
PPM_NLS	0.2020	1	136	3	0.1728	2	62	2	3	5	8
PPM_ML	0.3479	5	308	6	0.2240	4	136	5	9	11	20
PPM_Mixed	0.2774	4	153	5	0.3238	6	89	4	10	9	19
Empirical Percentile Model	0.2284	2	138	4	0.1922	3	72	3	5	7	12

* Abbreviations are as introduced in the text.

observed number of trees per 1000 m^2 within each diameter class (Reynolds et al. 1988, Liu et al. 2004), and indicates the typical misfit of a model, because the EI will be small only when the model predicts well for all diameter classes. Kolmogorov-Smirnov and EI test statistics were determined for both the modeling and the validation data sets, and further compared by Wilcoxon signed-ranks test in order to find statistically significant differences between the examined modeling approaches, regarding their goodness of fit. Different ranks were assigned to the methods distinguished as significantly different by the Wilcoxon test, for each test statistic and data set separately, and the overall method ranking was based on the rank totals summed from the 4 respective groups.

Results and Discussion

The final regressions required for application of the different modeling methods, their coefficients of determination (R^2), and their Root Mean Square Errors (RMSE) are presented in Table 2. In the PPM_Mixed method, both regression coefficients were initially represented by a fixed and a random part, but the random part of the shape parameter could not be regressed satisfactorily on the stand level variables and for this reason the model was reformulated with a mixed-scale and fixed-shape parameter (Table 2, Eq. 8). Two of the examined methods for parameter estimation (PRM_S and PPM_NLS) modeled successfully the pattern of all distributions in both the modeling and the validation data sets, which was revealed by the K-S test. The Empirical Percentile Model and the PRM_L method failed to model 2 and 3 of the 402 distributions, respectively. The PRM_B and PPM_Mixed methods, which modeled acceptably the distributions of the modeling data set (failed in 3 and 7 cases, respectively), showed poor predictive ability for the validation data set (failed in 18 and 39 cases, respectively). The results from the comparison of the examined methods are shown in Table 3. The total sums of the ranks varied between 6 and 20 and the methods of best diameter distribution modeling potential were the two parameter recovery methods through moments (PRM_S and PRM_L) and the parameter prediction method by non-linear least squares estimation at the first step of its application (PPM_NLS).

The parameter recovery methods have generally proved better to the parameter prediction ones in estimation of the Weibull frequency distribution function (Liu et al. 2004). The function relating the quadratic mean to the arithmetic mean diameter used in PRM_S (Eq. 5) is theoretically preferred, because it ensures that the predicted values for the arithmetic mean diameter will not exceed the quadratic mean and takes into account the influence of the stand variables. On the other hand, the simple linear regression in the method PRM_L estimated in the present study intrinsically overcomes the problem of higher arithmetic mean diameter (Table 2) and allows modeling the diameter distribution only through the value of the quadratic mean diameter. The parameter prediction approach PPM_NLS also showed very good results in the present study, which can be largely accounted to the very strong relationships of the function parameters to the stand level variables. PPM_NLS appeared much better than PPM_MLE, although the maximum-likelihood estimation (MLE) has been preferred as the one providing unequivocally the best possible distribution. The advantage, however, of the MLE is more pronounced when the primary purpose of its application is the distribution estimation itself and is recommended for large samples. Cao et al. (2004), on the other hand, examined different ways for estimating the 3 parameters of the Weibull function involving different prediction methods and combinations with recovery approaches. The study suggested that the limitations of the parameter prediction approaches can be successfully overcome by reformulating the optimization criterion for parameter estimation in two ways: maximizing the sum of the log-likelihood values or minimizing the error sum of squares of the cumulative distribution functions from all plots. The parameter recovery method (PRM_B) proposed by Cao et al. (1982) has been designed to practically utilize the information about the stand basal area per hectare, easily obtained through relascopic sample plots. Fitting Eq. 7, however, requires exhaustive, often difficult to converge calculations. In spite of the theoretically logical derivation of this method, its poor goodness of fit and estimation efforts make it disadvantageous in comparison to the other tested parameter recovery methods. Mixed-effect models, which have been successfully developed and applied as an alternative to the purely deterministic models, are preferable mainly when calibration through supplementary observations of the dependent variable is involved. Such calibration in the diameter distribution modeling is barely justified because an additional determination of the frequency of even a single diameter size class requires many additional measurements, which makes such approach meaningless. Prediction of the random parameters through the stand level variables was attempted in this study, but the simplified mixed-model technique proposed here did not produce satisfactory results. The poor predictability of the PPM_Mixed for the validation data set and the lack of significant relationship of the random component of the shape parameter to the stand level variables suggest that this approach has to be abandoned in modeling diameter distributions. A case study on *Pinus taeda* plantations (Borders and Patterson 1990) showed that an empirical percentile function can exceed the goodness of fit of the Weibull frequency function mainly because of better modeling of the non-unimodal distributions. This method did not show here superiority to the Weibull function estimated by the parameter recovery through moments. Although not connected to a particular known function, the percentile model outlines through its 12 percentiles a specific distribution pattern, which should be characteristic to the investigated type of stands. Thus, the distribution pattern estimated by this method did not represent the diameter distributions of the Scots pine plantations better than the Weibull frequency distribution function.

Acknowledgements

This article represents part of the research work under the implementation of Marie Curie Intra-European Fellowship Project PIEF-GA-2009-235039/25.08.2009.

References

Bailey R.L., Dell T.R. 1973. Quantifying diameter distributions with the Weibull function. Forest Science 19: 97–104.

Borders B.E., Patterson W.D. 1990. Projecting Stand Tables: A Comparison of the Weibull Diameter Distribution Method, a Percentile-Based Projection Method, and a Basal Area Growth Projection Method. Forest Science 36: 413–424.

Borders B.E., Souter R.A., Bailey R.L., Ware K.D. 1987. Percentile-Based Distributions Characterize Forest Stand Tables. Forest Science 33: 570–576.

Cao Q.V. 2004. Predicting parameters of a Weibull function for modeling diameter distribution. Forest Science 50: 682–685.

Cao Q.V., Burkhart H.E., Lemin R.C. 1982. Diameter distributions and yields of thinned loblolly pine plantations. School of Forestry and Wildlife Resources Publication No FWS-1-82, Virginia Polytechnic Institute and State University Blacksburg, Virginia 24061. 62 p.

Diéguez-Aranda U., Castedo-Dorado F., Álvarez-González J.G., Rodríguez-Soalleiro R. 2006. Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). Ecological Modeling 191: 225–242.

Efremov R. 2006. Effect of selection and linear-selection thinnings on the diameter growth of *Pinus silvestris* L. plantations. Sofia. Forest Science 3: 47–61. (in Bulgarian).

Krastanov K., Belyakov P., Shikov K. 1980. Dependencies in the structure, growth and productivity of the Scots pine plantations and thinning activities in them. Research report. Forest Research Institute of BAS, Sofia (in Bulgarian).

Little S.N. 1983. Weibull diameter distributions for mixed stands of western conifers. Canadian Journal of Forest Research 13: 85–88.

Liu C., Zhang S.Y., Lei Y., Newton P.F., Zhang L. 2004. Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. Canadian Journal of Forest Research 34: 2424–2432.

Marinov I. 2002. Stationary studies of hydrological and erosion processes in forest ecosystems. Research report for the period 2000-2002, Forest Research Institute of BAS, Sofia (in Bulgarian).

Marinov I., 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 a 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. European Journal of Forest Research 125: 427–439.

Mihov I.V. 1986. A method for determination of the type of growth of forest stands and its application to site index determination of Scots pine stands, DSc thesis, University of Forestry, Sofia (in Bulgarian).

Reynolds M.R., Burk T.E., Huang W.-C. 1988. Goodness-of fit tests and model selection procedures for diameter distribution models. Forest Science 34 (2): 373–399.

Siipilehto J. 2009. Modelling stand structure in young Scots pine dominated stands. Forest Ecology and Management, 257 (1): 223–232.

Stankova T., Zlatanov Tz. 2010. Modeling diameter distribution of Austrian black pine (*Pinus nigra* Arn.) plantations: a comparison of the Weibull frequency distribution function and percentile-based projection methods. European Journal of Forest Research, 129 (6): 1169–1179.

Stankova T., Shibuya M., Hagihara A. 2002. A method for density control of forest plantations. Sofia, Forest Science 2: 27–38.

Stankova T., Stankov H., Shibuya M. 2006. Mean-dominant height relationships for Scotch pine and Austrian black pine plantations Bulgaria. Ecological Engineering and Environmental Protection 2: 59–66.