JOURNAL OF FOREST SCIENCE, *54*, 2008 (12): 566–571

Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*)

Y. Lei

Research Institute of Resource Information and Techniques, Chinese Academy of Forestry, Beijing, China

ABSTRACT: Weibull distribution was used to fit tree diameter data collected from 86 sample plots located in Chinese pine stand in Beijing. To estimate the Weibull distribution parameters, three methods [namely maximum likelihood estimation method (MLE), method of moment (MOM) and least-squares regression method (LSM)] were compared and evaluated on the basis of the mean square error (MSE) and sample size. For these sample plots, the moment method was superior for estimating the parameters of Weibull distribution for tree diameter distribution.

Keywords: Weibull distribution; diameter distribution; parameter estimation

Tree diameter distributions play an important role in stand modelling. A number of different distribution functions have been used to model diameter distributions, including Beta, Lognormal, Johnson's Sb, and Weibull ones. The Weibull distribution, introduced by Bailey and Dell (1973) as a model for diameter distributions, has been applied extensively in forestry due to (1) its ability to describe a wide range of unimodal distributions including reversed-J shaped, exponential, and normal frequency distributions, (2) the relative simplicity of parameter estimation, and (3) its closed cumulative density functional form (e.g. Bailey, Dell 1973; Schreuder, Swank 1974; Schreuder et al. 1979; Little 1983; Rennolls et al. 1985; Mabvurira et al. 2002), and (4) its previous success in describing diameter frequency distributions within boreal stand types (e.g. Bailey, Dell 1973; Little 1983; Kilkki et al. 1989; Liu et al. 2004; Newton et al. 2004, 2005).

It is important that different estimation methods are compared to fit parameters of the Weibull probability density function (PDF) from given tree diameter breast height (dbh) data in forest inventory because the estimate parameters play a major role in developing a stand-level diameter distribution yield model based on stand variables employing the parameter prediction method, i.e. expressing the parameters of a probability density function (PDF) characterizing the diameter frequency distribution as a function of stand-level variables (Hyink, Moser 1983). Therefore, many other methods have been proposed to estimate the parameters of Weibull PDF distribution in forestry, such as the maximum likelihood estimation (MLE), the percentile estimation (PCT), and the method of moment (MOM) estimation. MLE is generally considered the best as it is asymptotically the most efficient method, and thus it is the most frequently used method to estimate parameters of distributions. However, the MLE does not exist in cases where the likelihood function can be made arbitrarily large. This occurs, for example, to distributions whose range depends on their parameters, such as the three-parameter Weibull distribution as we found in our simulation study. Some other methods have been proposed to estimate the parameters of the Weibull distribution, such as the ME, the PCT and the least-squares method (LSM). Zarnoch and Dell (1985) com-

The author is very grateful to MOST for its support of this work through Project 2006BAD23B02 and to the Inventory Institute of Beijing Forestry for its data.

pared the Weibull distribution estimation methods of both PCT and MLE based on the mean square error (MSE) in which there is a difference between the estimate and the true value of the parameter. They found that the MLE is superior in accuracy and has a smaller MSE compared with the PCT. Shiver (1988) evaluated three-parameter estimate methods (MLE, PCT and MOM) of the Weibull distribution in unthinned slash pine plantations based on the MSE and the conclusion supports the results of Zarnoch and Dell (1985).

The LSM has consistently been found to be superior for estimating the parameters of Sb distribution (Zhou, McTague 1996; Kamziah et al. 1999; Zhang et al. 2003) in forestry applications, but the LSM is used very little for estimating the parameters of Weibull distribution in forestry applications. The LSM provides alternatives to the MLE and MOM. Additionally, this method has an advantage in computation that most of the statistical software packages currently available (S-Plus, SAS, SPSS, ...) support the least-squares estimation but may not support the MLE and MOM, therefore it is worth introducing the LSM for fitting the Weibull distribution and comparing their performances with the MLE and MOM.

The objective of this research is to assess and compare the accuracy of the above three estimators of two-parameter Weibull distribution. Computer simulation techniques are used to generate Weibull populations with known parameters and the estimators are analyzed and evaluated from Chinese pine (*Pinus tabulaeformis*) data and simulation data using appropriate statistical procedures.

MATERIALS AND METHODS

Field data description

The data were provided by the Inventory Institute of Beijing Forestry. They consist of a systematic sample of permanent plots with a 5-year re-measurement interval. From the inventory plots over the whole of Beijing, all plots with 10 trees at least were used in this study (see Table 1), i.e. eighty-six 0.067 ha permanent sample plots (PSPs) located in plantations situated on upland sites throughout north-western Beijing. The PSPs data consisted of 256 measurements obtained in the following years: 1987, 1991, 1996 and 2001. All 256 measurement data of 86 sample plots were selected to estimate the two-parameter Weibull function using MLE, MOM and LSM methods in order to consistently compare the three different estimators.

Methods of estimation

The probability and cumulative distribution functions of the three-parameter Weibull distribution for a random variable D are

$$f(D;a,b,c) = \frac{c}{b}\left(\frac{D-a}{b}\right)^{c-1} \exp\left(-\left(\frac{D-a}{b}\right)^{c}\right) = 0 \quad (a \le D \le \infty) \quad (D < a) \tag{1}$$

$$F(D;a,b,c) = 1 - \exp\left(-\left(\frac{D-a}{b}\right)^{c}\right) \tag{2}$$

where:
D – diameter at breast height (in cm),
a – location parameter,
b – scale parameter,
c – shape parameter.

The parameters of Equation (1) were estimated from the individual tree diameter data of each set of diameter data by maximum likelihood estimation. In some plots the procedure of maximum likelihood estimates can result in a negative value for the location parameter a. The parameter a can be considered as the smallest possible diameter in the stand and thus it should be between 0 and the mini-

Table 1. Descriptive statistics of stand and tree variables

	Stand variable (86 plots)					Tree variable (n = 15,676 trees)	
	dbh (cm)	age (years)	N (trees/ha)	H (m)	BA (m^2/ha)	dbh (cm)	BA (m^2/tree)
Mean	10.53	28.5	918	6.22	8.71	10.25	0.00953
Standard deviation	2.91	8.74	540	2.47	6.13	4.06	0.00821
Min.	5.82	11.00	150	2.50	0.45	0.50	0.00196
Max.	21.92	53.00	2,354	19.50	33.50	36.80	0.10631

dbh – diameter at breast height; N – stand density; H – average height of dominant and codominant trees; BA – basal area; Mean, Min., Max. – mean, minimum and maximum diameter at breast height respectively

mum observed value in some cases (Bailey, Dell 1973). An approximation to this smallest possible diameter is given by minimum diameter at breast height (Dmim), which is the minimum observed diameter on the sample plots. By arbitrarily setting *a* to 0.5 Dmim in some studies and then estimating parameters *b* and *c*, three-parameter Weibull function can be obtained (Kilkki et al. 1989). Thus, the two-parameter Weibull distribution was considered in this study as follows

$$F(D;b,c) = 1 - \exp\left(-\left(\frac{D}{b}\right)^c\right) \quad (3)$$

Three methods (MLE, MOM and LSM) mentioned above were used to estimate the Weibull distribution in this study.

Maximum likelihood estimator (MLE)

The method of maximum likelihood is a commonly used procedure for the Weibull distribution in forestry because it has very desirable properties. Estimation of the parameters by maximum likelihood has been found to produce consistently better goodness-of-fit statistics compared to the previous methods, but it also puts the greatest demands on the computational resources (Cao, McCarty 2005). Consider the Weibull PDF given in (1), then the likelihood function (*L*) will be

$$L(D_1, ..., D_n;b,c) = \prod_{i=1}^{n}\frac{c}{b}\left(\frac{D}{b}\right)^{c-1}\exp\left(-\left(\frac{D}{b}\right)^c\right) \quad (4)$$

On taking the logarithms of (4), differentiating with respect to *b* and *c* respectively, and satisfying the equations

$$b = \left[n^{-1}\sum_{i=1}^{n}D_i^c\right]^{1/c} \quad (5)$$

$$c = \left[\left(\sum_{i=1}^{n}D_i^c \ln D_i\right)\left(\sum_{i=1}^{n}D_i^c\right)^{-1} - n^{-1}\sum_{i=1}^{n}\ln D_i\right]^{-1} \quad (6)$$

The value of *c* has to be obtained from (6) by the use of standard iterative procedures (i.e. Newton-Raphson method) and then used in (5) to obtain *b*.

Methods of moments (MOM)

The method of moments is another technique commonly used in the field of parameter estimation. In the Weibull distribution, the *k* moment readily follows from (1) as

$$m_k = \left(\frac{1}{b}\right)^{k/c}\Gamma\left(1 + \frac{k}{c}\right) \quad (7)$$

where:
Γ – gamma function, $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$, $(s > 0)$.

Then from (7), we can find the first and the second moment as follows

$$m_1 = \mu = \left(\frac{1}{b}\right)^{1/c}\Gamma\left(1 + \frac{1}{c}\right) \quad (8)$$

$$m_2 = \mu^2 + \sigma^2 = \left(\frac{1}{b}\right)^{2/c}\left\{\Gamma\left(1 + \frac{2}{c}\right) - \left[\Gamma\left(1 + \frac{1}{c}\right)\right]^2\right\} \quad (9)$$

where:
σ^2 – variance of tree diameters in a plot,
m_1, m_2 – arithmetic mean diameter and quadratic mean diameter in a plot, respectively.

When m_2 is divided by the square of m_1, the expression of obtaining *c* only is

$$\frac{\sigma^2}{\mu^2} = \frac{\Gamma(1 + \frac{2}{c}) - \Gamma^2(1 + \frac{1}{c})}{\Gamma^2(1 + \frac{1}{c})} \quad (10)$$

On taking the square roots of (10), the coefficient of variation (*CV*) is

$$CV = \frac{\sqrt{\Gamma(1 + \frac{2}{c}) - \Gamma^2(1 + \frac{1}{c})}}{\Gamma^2(1 + \frac{1}{c})} \quad (11)$$

In order to estimate *b* and *c*, we need to calculate the *CV* of tree diameters in plots and get the estimator of *c* in (11). The scale parameter (*b*) can then be estimated using the following equation

$$\hat{b} = \{\overline{x} / \Gamma[(1/\hat{c}) + 1]\}^{\hat{c}} \quad (12)$$

where:
$\overline{x}$ – mean of the tree diameters.

Least squares method (LSM)

For the estimation of Weibull parameters, the least-squares method (LSM) is extensively used in engineering and mathematics problems. We can get a linear relation between the two parameters taking the logarithms of (3) as follows

$$\ln\ln\left[\frac{1}{1 - F(D)}\right] = c\ln D - c\ln b \quad (13)$$

where:
$Y = \ln\{-\ln[1-F(D)]\}$
$X_i = \ln D$
$\lambda = -c\ln b$.

Let $D_1, D_2, ..., D_n$ be a random sample of *D* and *F*(*D*) is estimated and replaced by the median rank method as follows:

$$F(D) = (i - 0.3)/(n + 0.4)\ (D_i, i = 1, 2, ..., n \text{ and } D_1 < D_2 < ... < D_n) \quad (14)$$

Table 2. Number of times minimizing MSE for 256 diameter frequency distribution measurements by method

Method	No. of times the method gives the best estimate	Mean	SD
MOM	152	2.7×10^{-3}	2.4×10^{-3}
MLE	35	5.3×10^{-3}	4.6×10^{-3}
LSM	69	3.84×10^{-3}	4.8×10^{-3}

because $F(D)$ of the mean rank method

$$[F(D) = i/(n + 1)]$$

may be a larger value for smaller i and a smaller value for larger i.

Thus, Equation (13) is a linear equation and is expressed as

$$Y = cX + \lambda \tag{15}$$

Computing c and λ by simple linear regression in (15) and the parameters c and b can be estimated as:

$$c = [\sum_{i}^{n} XY - 1/n(\sum_{i}^{n} X \sum_{i}^{n} Y]/[\sum_{i}^{n} X^2 - 1/n(\sum_{i}^{n} X)^2] \tag{16}$$

$$\lambda = 1/n(\sum_{i}^{n} Y - c/n \sum_{i}^{n} X \tag{17}$$

$$b = \exp(-\lambda/c) \tag{18}$$

Statistical criteria

For quantitative comparison of different estimators, mean square error (MSE) was used to test the estimators of the three methods by the 256 diameter frequency distribution measurements (observations) from 86 sample plots for field data in this study. MSE is a measure of the accuracy of the estimator. MSE can be calculated as below

$$MSE = \sum_{i}^{n} \left\{ \hat{F}(D_i) - F(D_i) \right\}^2 \tag{19}$$

where:

$\hat{F}(D_i) = 1 - \exp(-D_i/\hat{b})^{\hat{c}}$ – value of the cumulative distribution function (CDF) of the Weibull distribution evaluated at dbh of tree i in a plot by using different estimations,

$F(D_i)$ – observed cumulative probability of tree i in a plot,

n – number of trees in a plot.

In this study, testing and evaluation computations were completed using the Forstat statistical package (Tang et al. 2006).

RESULTS AND DISCUSSION

The 256 diameter frequency distribution measurements (observations) from 86 sample plots were used to estimate the two-parameter Weibull function based on the MLE, LSM and MOM. The best estimated method was evaluated according to minimum MSE, mean and SD MSE. Table 2 displays the summaries of the MSE indicator for 256 diameter frequency distribution measurements. From Table 2, the MOM produced the best estimate 152 times out of 256 diameter frequency distribution measurements, which is approximately 59.3%, followed by the LSM 69 times (27.0%) and the MLE 35 times (13.7%), respectively. The mean MSEs from 152 times in MOM, 69 times in LSM and 35 times in MLE are 2.7×10^3, 3.84×10^3 and 5.3×10^3 cm, respectively. The Weibull parameters c and b were estimated by the maximum likelihood method (MLE) for 35 diameter frequency distribution measurements. The parameter values of the MLE ranged as follows: $2.85 \leq c \leq 7.47$, $62.21 \leq b \leq 224.52$; the LSM for 69 diameter frequency distribution measurements, the parameter values ranged as follows: $2.45 \leq c \leq 10.69$, $66.20 \leq b \leq 186.51$; the MOM for 152 diameter frequency distribution measurements, the parameter values ranged as follows: $1.60 \leq c \leq 7.2$, $63.54 \leq b \leq 241.27$. The MOM achieved good estimated results because it involved more calculations and required more computation time than the LSM or the MLE (Al-Fawzan 2000). Although the results from the LSM and the MLS estimated methods were inferior to the MOM based on the MSE criterion in this study, the LSM and the MLE aimed at fitting the entire diameter distribution itself (rather than just the average diameter or plot-level diameter attributed such as diameter moments). Therefore, it seemed reasonable to expect the LSM or the MLE method in estimating the Weibull distribution function. Actually, Cao and McCarty (2005) reported that the cumulative distribution function (CDF) regression method produced better results than those from the MOM based on the chi-square statistic for loblolly

pine plantations in the southern United States because the CDF regression aimed at fitting the CDF of diameter distribution. Also, the LSM improves the fitting of the distribution because more information is used than in the MOM.

CONCLUSION

In this study, the good results of the MOM in terms of the number of times for the lowest values of MSE indicated that the MOM was a superior method to estimate the diameter distribution of Weibull function for Chinese pine stand in Beijing. However, from the aspect of estimated performance, the LSM and the MLE of fitting Weibull function were also good methods because the methods are easy and quick estimates well as there exists a lot of software to estimate the parameters of Weibull distribution. Specially, the LSM method improves the fitting of tree diameter distributions because more information is used than in the MOM method. Since the regression method uses simple linear regression to estimate the parameters *c* and *b* of the Weibull function, it may be an appropriate method for predicting a future stand.

Acknowledgements

The author would like to thank Dr. Mohammad Al-Fawzan for providing his information to this paper.

References

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

CAO Q.V., McCARTY S.M., 2005. Presented at the Thirteenth Biennial Southern Silvicultural Research Conference. Memphis, TN.

HYINK D.M., MOSER J.W., 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science, *29*: 85–95.

KAMZIAH A.K., AHMAD M.I., LAPONGAN J., 1999. Nonlinear regression approach to estimating Johnson SB parameters for diameter data. Canadian Journal of Forest Research, *29*: 310–314.

KILKKI P., MALTAMO M., MYKKANEN R., PAIVINEN R., 1989. Use of the Weibull function in estimating the basal area dbh-distribution. Silva Fennica, *23*: 311–318.

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

LIU C.M., ZHANG S.Y., LEI Y., ZHANG L.J., 2004. Comparison of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in eastern Canada. Canadian Journal of Forest Research, *34*: 2424–2432.

MABVURIRA D., MALTAMO M., KANGAS A., 2002. Predicting and calibrating diameter distributions of *Eucalyptus grandis* (Hill) Maiden plantations in Zimbabwe. New Forests, *23*: 207–223.

NEWTON P.F., LEI Y., ZHANG S.Y., 2004. A parameter recovery model for estimating black spruce diameter distribution within the context of a stand density management diagram. The Forestry Chronicle, *3*: 349–358.

NEWTON P.F., LEI Y., ZHANG S.Y., 2005. Stand-level distance-independent diameter distribution model for black spruce plantations. Forest Ecology and Management, *209*: 181–192.

RENNOLLS K., GEARY D.N., ROLLINSON T.J.D., 1985. Characterizing diameter distributions by the use of the Weibull distribution. Forestry, *58*: 57–66.

SCHREUDER H.T., SWANK W.T., 1974. Coniferous stands characterized with the Weibull distribution. Canadian Journal of Forest Research, *4*: 518–523.

SCHREUDER H.T., HAFLEY W.L., BENNETT F.A., 1979. Yield prediction for unthinned natural slash pine stands. Forest Science, *25*: 25–30.

SHIVER B.D., 1988. Sample size and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distribution. Forest Science, *34*: 809–814.

TANG S., LAN K.J., LI Y., 2006. Guide of ForStat.2.0. (Unpublish.)

ZARNOCH S.J., DELL T.R., 1985. An evaluation of percentile and maximum likelihood estimators of Weibull parameters. Forest Science, *31*: 260–268.

ZHANG L., PACKARD K.C., LIU C., 2003. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce-fir stands in northeastern North America. Canadian Journal of Forest Research, *33*: 1340–1347.

ZHOU BAILIN, McTAGUE J.P., 1996. Comparison and evaluation of five method of estimation of the Johnson system parameters. Canadian Journal of Forest Research, *26*: 928–936.

AL-FAWZAN MOHAMMAD, 2000. Method for estimating the parameters of the Weibull distribution. (Unpublish.)

Received for publication July 8, 2008
Accepted after corrections September 1, 2008

Hodnotenie troch metód na určenie parametrov Weibullového rozdelenia čínskej borovice (*Pinus tabulaeformis*)

ABSTRAKT: Na vyrovnanie hrúbok stromov zozbieraných z 86 výskumných plôch čínskej borovice v Pekingu sa použilo Weibullove rozdelenie. Pri určovaní parametrov Weibullového rozdelenia sa prostredníctvom strednej kvadratickej chyby a počtu prípadov porovnávali a hodnotili tri metódy, menovite metóda maximálnej vierohodnosti – MLE, momentová metóda – MOM a regresná metóda najmenších štvorcov LSM. Na určenie parametrov Weibullového rozdelenia hrúbok stromov výberových plôch bola najlepšia momentová metóda.

Kľúčové slová: Weibullove rozdelenie; rozdelenie hrúbok; určenie parametrov

Corresponding author:

Prof. Dr. Yuancai Lei, Research Institute of Resource Information and Techniques, Chinese Academy of Forestry, Beijing 100091, China, P. R.
tel.: + 010 6288 9199, fax: + 010 6288 8315, e-mail: yclei@caf.ac.cn, leiycai@yahoo.com