NEW METHODS FOR ESTIMATING PARAMETERS OF WEIBULL FUNCTIONS TO CHARACTERIZE FUTURE DIAMETER DISTRIBUTIONS IN FOREST STANDS

Quang V. Cao and Shanna M. McCarty[1]

Abstract—Diameter distributions in a forest stand have been successfully characterized by use of the Weibull function. Of special interest are cases where parameters of a Weibull distribution that models a future stand are predicted, either directly or indirectly, from current stand density and dominant height. This study evaluated four methods of predicting the Weibull parameters, using data from loblolly pine (*Pinus taeda* L.) plantations. One method, the CDF Regression, consistently produced better goodness-of-fit statistics than those from the other methods.

INTRODUCTION

The Weibull function was introduced by Bailey and Dell (1973) to model diameter distributions in forest stands. The advantages of the Weibull include its flexibility to fit shapes commonly found in both uneven-aged and even-aged stands and the ease of computing probabilities (or proportions of trees in diameter classes) without the need for numerical integration. Cao (2004) developed a new approach, termed CDF Regression, that outperformed the existing methods in estimating Weibull parameters of a stand from known density and dominant height. The objective of this study was to evaluate four methods to predict the parameters of Weibull functions that modeled diameter distributions of a future stand, whose attributes need to be predicted from current stand attributes.

DATA

Data were from the Southwide Seed Source Study, which involved 15 loblolly pine (*Pinus taeda* L.) seed sources planted at 13 locations across 10 Southern States (Wells and Wakeley 1966). Seedlings were planted at a 6 foot x 6 foot spacing. Each 0.04-acre plot consisted of 49 trees measured 4 times at ages 10, 15 or 16, 20 or 22, and 25 or 27. A subset (100 plots) of the original data was randomly selected as the fit data set, to be used for fitting the models. Furthermore, only one growing period was randomly chosen from each plot. The fit data set therefore contained 100 growth periods. Another 100 plots were randomly selected from the remaining original data to form a validation data set. All 3 growing periods from these plots were used to evaluate the methods, resulting in a total of 300 observations in the evaluation data set. Table 1 shows summary statistics for stand attributes at the end of each growth period for the fit and validation datasets.

Table 1—Means (and standard deviations) of stand attributes at the end of each growth period for the fit and validation data

Age	n	Dominant height	Trees per acre	Basal area *sq.ft./acre*
- - - - - -	- - - - - -	Fit data	- - - - - -	- - - - - -
15	30	42 (4)	778 (212)	146 (37)
16	4	55 (3)	537 (127)	138 (18)
20	33	58 (7)	484 (117)	148 (31)
25	20	61 (6)	436 (171)	139 (47)
27	13	63 (10)	327 (90)	129 (38)
- - - - - -	- - - - - -	Validation data	- - - - - -	- - - - - -
15	86	43 (6)	702 (187)	134 (27)
16	14	53 (4)	515 (79)	130 (14)
20	95	56 (7)	526 (131)	153 (34)
22	5	46 (7)	385 (48)	83 (36)
25	72	65 (8)	489 (133)	176 (41)
27	28	67 (10)	322 (81)	141 (30)

METHODS

The Weibull probability density function (pdf), used in this study to characterize diameter distribution, has the following form:

$$f(x) = \left(\frac{c}{b}\right)\left(\frac{x-a}{b}\right)^{c-1} \exp\left[-\left(\frac{x-a}{b}\right)^{c}\right] \qquad (1)$$

where a, b, and c are the location, scale, and shape parameters, respectively, and x is tree diameter at breast height (d.b.h.).

The annual growth model was employed in this study to project stand attributes, because it was found by Ochi and Cao (2003) to provide better stand predictions than compatible growth models. The stand attributes at time (t+1) were projected from time t as follows:

$$\hat{N}_{t+1} = \hat{N}_t^{A_t / A_{t+1}}[b_1 + b_2 / A_t + b_3\hat{N}_t]^{(1-A_t / A_{t+1})} \qquad (2)$$

$$\hat{B}_{t+1} = \hat{B}_t^{A_t / A_{t+1}}[b_1 + b_2 / \hat{H}_t + b_3 / \ln(\hat{B}_t)]^{(1-A_t / A_{t+1})} \qquad (3)$$

$$\hat{D}\min_{t+1} = \hat{D}\min_t^{A_t / A_{t+1}}[b_1 + b_2 / A_t]^{(1-A_t / A_{t+1})} \qquad (4)$$

$$SD_{t+1} = SD_t^{A_t / A_{t+1}}[b_1 + b_2 / \ln(\hat{N}_t)]^{(1-A_t / A_{t+1})} \qquad (5)$$

[1] Professor and Gilbert Fellow, respectively, Louisiana State University AgCenter, School of Renewable Natural Resources, Baton Rouge, LA 70803.

Citation for proceedings: Connor, Kristina F., ed. 2006. Proceedings of the 13th biennial southern silvicultural research conference. Gen. Tech. Rep. SRS–92. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 640 p.

where A = stand age in years, $\hat{H}$ = the predicted dominant height (average height of the dominants and codominants), $\hat{N}$ = predicted number of trees per acre, $\hat{B}$ = the predicted stand basal area in square feet per acre, $\hat{D}$min = the minimum diameter in inches, SD = the predicted standard deviation of diameter, and the subscripts place the attributes at time t or (t+1). In addition, the b_i's are regression coefficients.

The future 93rd diameter percentile ($\hat{D}93_2$) at the end of the growth period (age A_2) was predicted from the future quadratic mean diameter ($\hat{D}q_2$) and relative spacing (RS_2):

$$\hat{D}93_2 = \hat{D}q_2[1 + \exp(b_1 + b_2 RS_2)] \tag{6}$$

where $\hat{D}q_2 = \sqrt{\hat{B}_2 / 0.005454}$, $RS_2 = \sqrt{43560 / \hat{N}_2} \ / \ \hat{H}_2$, and $\hat{H}_2$, $\hat{N}_2$, and $\hat{B}_2$ are dominant height, number of trees, and basal area per acre at age A_2, respectively.

Borders (1989) method was used to simultaneously estimate regression parameters of equations to predict N_2, B_2, $D\text{min}_2$, SD_2, and $D93_2$.

The following methods to obtain the Weibull parameters were evaluated.

Method 1—Moment Estimation

The Weibull location parameter (a) must be smaller than the predicted minimum diameter in the stand ($\hat{D}\text{min}_2$). We set a = 0.5$\hat{D}\text{min}_2$ since Frazier (1981) found that this gave best results in terms of goodness-of-fit. The other Weibull parameters, b and c, were recovered from the first two moments of the diameter distribution (Cao and others 1982):

$$b = (\hat{\hat{D}}_2 - a) / \Gamma_1 \tag{7}$$

$$\hat{D}q_2^2 + a^2 - 2a\hat{\hat{D}}_2 - b^2\,\Gamma_2 = 0 \tag{8}$$

where $\hat{\hat{D}}_2 = \sqrt{\hat{D}q_2^2 - SD_2^2}$ and $\Gamma_i = \Gamma(1 + i/c)$.

Method 2—Hybrid

The Weibull location parameter was also computed from a = 0.5$\hat{D}\text{min}_2$. In the hybrid method, developed by Baldwin and Feduccia (1987), the b and c parameters were recovered from a moment (the quadratic mean diameter or $\hat{D}q_2$) and a percentile (the 93rd diameter percentile or $\hat{D}93_2$):

$$b = (\hat{D}93_2 - a) / 2.659261/c \tag{9}$$

$$0 = a2 - \hat{D}q_2 + 2a\,(\hat{D}93_2 - a)\,\Gamma_1 / 2.65926^{1/c} + a\,(\hat{D}93_2 - a)2\,\Gamma_2 / 2.65926^{2/c} \tag{10}$$

where 2.65926 = –ln(1 – 0.93).

Method 3—CDF Regression

Similar to the previous two methods, the Weibull location parameter was set at a = 0.5$\hat{D}\text{min}_2$. The scale and shape parameters were computed from:

$$b = \frac{-a\Gamma_1 + \sqrt{a^2(\Gamma_1^2 - \Gamma_2) + Dq^2\Gamma_2}}{\Gamma_2} \tag{11}$$

$$c = \exp(b_1 + b_2 RS_2 + b_3\hat{H}_2) \tag{12}$$

As in the procedure developed by Cao (2004), this method iteratively searched for the coefficients b_i's in equation (12) to minimize the following function:

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(F_{ij} - \hat{F}_{ij})^2/n_i \tag{13}$$

where F_{ij} = observed cumulative probability of tree j in the i^{th} plot, $\hat{F}_{ij} = 1 - \exp\{-[(x_{ij} - a)/b]^c\}$ = value of the cumulative distribution function (cdf) of the Weibull distribution evaluated at x_{ij}, x_{ij} = d.b.h. of tree j in the i^{th} plot, n_i = number of trees in the i^{th} plot, and p = number of plots. This approach, similar to the procedure developed by Cao (2004), fit a regression equation for c, but the objective was to minimize the sum of squares of error with respect to the cdf (hence the name CDF Regression), rather than to c.

Method 4—Cumulative TPA Regression

This method is similar to method 3 (CDF Regression), except that equation (12) was replaced with:

$$c = \exp(b_1 + b_2 RS_2 + b_3/\hat{H}_2 + b_4\ln(\hat{N}_2)) \tag{14}$$

The coefficients b_i's in equation (14) were iteratively searched to minimize the following function:

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(T_{ij} - \hat{T}_{ij})^2/n_i \tag{15}$$

where T_{ij} and $\hat{T}_{ij}$ are observed and predicted cumulative number of trees per acre up to diameter x_{ij}.

Again, this approach was similar to fitting a regression equation for c, but the goal was to minimize the sum of squares of error with respect to the cumulative trees per acre (hence the name Cumulative TPA Regression), rather than to c.

EVALUATION

The following goodness-of-fit statistics were computed separately for each method and for each plot of the fit data and each plot-age combination of the validation data.

1. The chi-square statistic, and
2. Reynolds and others (1988) error indices for the i^{th} plot-age combination, based on number of trees per acre in each diameter class (EI_N) and basal area per acre in each diameter class (EI_B):

$$EI_{N,i} = \Sigma \mid m_{ik} - \hat{m}_{ik} \mid \tag{16}$$

$$EI_{B,i} = \Sigma \mid b_{ik} - \hat{b}_{ik} \mid \tag{17}$$

where m_{ik} and $\hat{m}_k$ = the observed and predicted number of trees per acre in diameter class k for the i^{th} plot-age combination, and b_{ik} and $\hat{b}_{ik}$ = the observed and predicted basal area per acre in diameter class k for the i^{th} plot-age combination. The sum includes all diameter classes in the i^{th} plot-age combination.

Table 2—Means[a] (and standard deviations) of goodness-of-fit statistics for the fit and validation data, by method

	Chi-square statistic		Error index based on TPA		Error index based on BA	
Method	Fit data	Validation data	Fit data	Validation data	Fit data	Validation data
Moment Estimation	15(29)	30(88)	*247*[b] (108)	267 (117)	68 (26)	75 (28)
Hybrid	8 (6)	11(25)	281 (132)	288 (111)	72 (28)	76 (29)
CDF Regression	*6* (5)	*7* (6)	250 (114)	*264* (113)	*66* (25)	*71* (28)
CTPA Regression	7 (5)	9(18)	265 (134)	276 (151)	69 (28)	73 (28)

[a] The smaller the statistic value, the better the fit.
[b] For each statistic, an underlined italic number indicates the best among four methods.

RESULTS AND DISCUSSION

Table 2 displays the summaries of the chi-square statistic and the error indices for both datasets. The CDF Regression method produced the lowest goodness-of-fit statistics for both fit and validation data in all but one case. The lone exception is its second rank in terms of EI_N for the fit dataset. On the other hand, the rankings of the rest of the methods were not clearcut, with each method received ranks varied from 2 to 4.

The same predicted variables ($\hat{N}_2$, $\hat{D}\text{min}_2$, and $\hat{D}q_2$) were employed in all methods. This means that all methods used the same location parameter and were constrained to produce the same total basal area per acre. The CDF Regression method performed well because the entire distribution (rather than just the average diameter or the 93rd diameter percentile) contributed to the fitting criterion.

Because total number of trees had to be predicted, error in the total number would in turn cause error in predicting number of trees in each diameter class. The Cumulative TPA Regression method was designed to optimize the Weibull parameter prediction by working directly with number of trees rather than proportions as did the CDF Regression. It was somewhat surprising to find that the CDF Regression was still the better method.

Cao (2004) concluded that the CDF Regression was best for predicting Weibull parameters to characterize current stands with known stand attributes. In this study, the CDF Regression was also found to be the most appropriate method for a future stand where its attributes had to be predicted from current attributes.

LITERATURE CITED

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

Baldwin, V.C., Jr.; Feduccia, D.P. 1987. Loblolly pine growth and yield prediction for managed West Gulf plantations. Res. Pap. SO-236. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 27 p.

Borders, B.E. 1989. Systems of equations in forest stand modeling. Forest Science. 35: 548-556.

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. Research Publication FWS-1-82. Blacksburg, VA: Virginia Polytechnic Institute and State University, School of Forestry and Wildlife. 62 p.

Frazier, J.R. 1981. Compatible whole-stand and diameter distribution models for loblolly pine plantations. Blackburg, VA: Virginia Polytechnic Institute and State University, School of Forestry and Wildlife. 125 p. Ph.D. dissertation.

Ochi, N.; Cao, Q.V. 2003. A comparison of compatible and annual growth models. Forest Science. 49: 285-290.

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: 373-399.

Wells, O.O.; Wakeley, P.C. 1966. Geographic variation in survival, growth, and fusiform-rust infection of planted loblolly pine. Forest Science Monograph 11. 40 p.