Quantifying Diameter Distributions with the Weibull Function

ROBERT L. BAILEY

T. R. DELL

Abstract. The Weibull probability density function is proposed as a diameter distribution model. Its advantages include flexibility in shape and simplicity of mathematical derivations. Estimation and interpretation of parameters are discussed and illustrated with published data. ***Forest Sci. 19:97–104.***

Additional key words. Exponential distribution, maximum likelihood, percentiles, point individual distance, simulation.

THE DISTRIBUTION OF DIAMETERS is the most potent simple factor for depicting the properties of a stand of trees. Diameter is generally well correlated with other important variables including volume, value, conversion cost, and product specifications. Quantification of the diameter distribution and its relationship to site, stand composition, age, and density is often valuable for both economic and biological purposes.

In this paper we consider the Weibull probability density function as a diameter distribution model. We discuss its properties, relate its advantages, and give techniques of parameter estimation. Several illustrative examples based on data taken from published works are presented.

Previous Models

In 1898 de Liocourt constructed a model based on the geometric progression for diameter distributions from uneven-aged forests (Meyer and Stevenson 1943). Meyer and Stevenson applied this general model, the exponential distribution, to forests of mixed species in Pennsylvania. Meyer (1952) and later Schmelz and Lindsey (1965) also found it satisfactory for diameter distributions. Leak (1965) gives a complete treatment of the application of this formula to reversed J-shaped diameter distributions.

Other systems and distributions which broaden the consideration to include the mound shape are the Gram-Charlier series (Meyer 1930), the Pearl-Reed growth curve (Nelson 1964, Osborne and Schumacher 1935), Pearsonian curves (Schnur 1934), the gamma distribution (Nelson 1964), and the three-parameter logarithmic-normal (Bliss and Reinker 1964). The beta distribution, which is essentially a reparameterization of Pearson's more general Type I, was applied to diameter distributions by Clutter and Bennett (1965). Applications of models based on the beta distribution were subsequently developed by McGee and Della-Bianca (1967) and Lenhart and Clutter (1971).

Choosing a Model

For consistency and simplicity it is desirable to select a single function capable of depicting the full range of unimodal, continuous shapes taken on by diameter distributions. Thus, the probability density function should cover the reversed J-shapes, plus mound shapes with varying degrees of either positive or negative skewness. Any constants in the model should be easily related to shape and location features of the distribution and thus vary in a consistent manner with stand characteris-

The authors are Mathematical Statisticians, Southern Forest Experiment Station, USDA Forest Service, New Orleans, La. Manuscript received July 3, 1972.

tics. One major application of the probability density function is integration to obtain proportions of the stand less than a stated diameter. The user should be able to calculate these values without laborious numerical integration or a computer. The function should provide a promising base for advanced developments. It should also be easily fitted to observed data using parameter estimators with good properties.

Certain previously used functions, particularly the beta, are flexible. However, varying degrees of difficulty in fitting and utilizing these functions have been encountered. The beta is fitted quite easily by moments, whereas fitting Gram-Charlier curves can be difficult. Generally, previous models have required numerical integration to determine the proportion of the total trees per acre by diameter classes and have parameters that are difficult to interpret. None of the functions previously considered have fully met the stated specifications.

Weibull Distribution

In a study of extreme values, Fisher and Tippett (1928) presented a new probability distribution which was independently derived by Weibull (1939) in studies of reliability of materials. Mann (1968) conjectured that emphasis on reliability analysis following World War II gave prominence to Weibull's work, and the distribution came to be associated with Weibull's name.

The probability density function (p.d.f.) of the two-parameter model for the Weibull random variable X, utilizing notation by Dubey (1967), is:

$$f(x) = (c/b)\,(x/b)^{c-1} \exp\{-(x/b)^c\}; \quad x \geqslant 0, b > 0, c > 0. \qquad (1)$$

It is characterized by the scale parameter b and the shape parameter c. In a more general formulation it also includes the location parameter a:

$$g(y) = (c/b)\,((y-a)/b)^{c-1} \exp\{-((y-a)/b)^c\}; \quad y \geqslant a, b > 0, c > 0. \qquad (2)$$

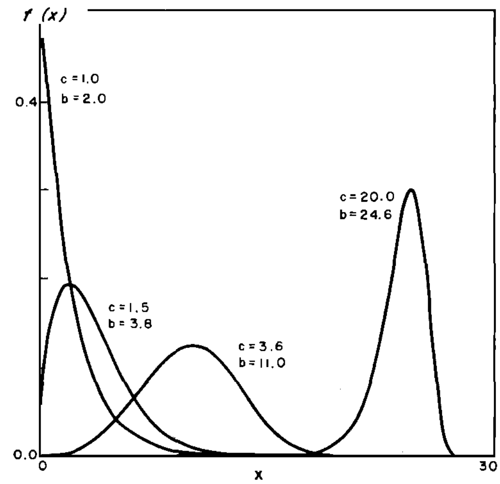

FIGURE 1. *Probability density function of the two-parameter Weibull distribution.*

These two models are related by the transformation $y = x + a$. Thus, no loss in generality results by restricting our discussion to equation (1). In a diameter distribution context the parameter $a \geqslant 0$ can be interpreted as the smallest possible diameter.

Motives for our consideration of this function in diameter distribution work are essentially those that guided the formulations of Weibull (1951). We present no biological basis for choosing a distribution function. A function that is simple and mathematically handy is specified, and its capabilities are explored. However, in one biological context the Weibull can be derived via assumptions and rationale. If trees are assumed to be distributed on a plane by a Poisson process as discussed by Pollard (1971), then the distribution of the distance from a randomly chosen point to the nearest individual is a Weibull with $a = 0$ and $c = 2$. This connection seems to have been missed by authors studying the problem of distance estimators of density.

The Weibull will assume the required variety of shapes (Fig. 1), and the shape is dependent on the value of c (Harter 1964). If $c < 1$, the curve is a reversed J. When $c = 1$, the exponential distribution results:

$$f(x) = (1/b) \exp\{-x/b\}; \quad 0 \leqslant x, b > 0. \tag{3}$$

For $1 < c < 3.6$, the density function is mound shaped and positively skewed. When $c = 2$, the Rayleigh, a special case of the χ distribution, results. With $c \simeq 3.6$, the Weibull approximates a normal distribution. As c is increased above 3.6 the distribution becomes progressively more negatively skewed, and as $c \rightarrow \infty$ it approaches a spike over a single point. Thus, estimated values of c should be directly related to age in the development of uncut even-aged stands for a given site. Values of $c \leqslant 1$ should occur in all-aged stands of tolerant species.

The cumulative distribution function of the Weibull, $F(x)$, is readily obtained by a change of variable technique.

$$\begin{aligned} F(x) &= \int_0^x (c/b)\,(t/b)^{c-1} \exp\{-(t/b)^c\}\, dt \\ &= \int_0^x \exp\{-(t/b)^c\}\, d[(t/b)^c] \\ &= \int_0^{(x/b)^c} \exp(-\mu)\, d\mu \\ &= 1 - \exp\{-(x/b)^c\}. \end{aligned} \tag{4}$$

This closed-form expression facilitates succeeding mathematical endeavors. With N stems-per-acre the number of trees in a diameter class having midpoint x and width $2w$ is given by

$$\begin{aligned} N_x &= N\{F(x+w) - F(x-w)\} \\ &= N\{e^{-[(x-w)/b]^c} - e^{-[(x+w)/b]^c}\}. \end{aligned}$$

Equation (4) can be solved for x,

$$x = b\{-\ln[1 - F(x)]\}^{1/c},$$

"ln" denoting the natural logarithm. Replacing $F(x)$ by a random variable with a uniform distribution on the interval (0,1) provides a mechanism for generating diameter distributions in simulation studies. Most computer systems make provision for generating this standardized uniform distribution. No previous model for diameter distributions leads to such efficient simulation.

If $F(x)$ is evaluated for the positive value $x = b$, the scale parameter is shown to be approximately equal to the 63rd percentile of the distribution.

$$\begin{aligned} F(b) &= 1 - \exp\{-(b/b)^c\} \\ &= 1 - e^{-1} \simeq 0.63 \end{aligned}$$

Thus, b can be interpreted as the diameter such that 63 percent of all the trees are smaller. However, as will be shown later, this is not the best approach for estimating b. In the case of equation (2), $(a + b)$ is the 63rd percentile.

Because the Weibull function has proved attractive in various areas of application and theory, there are many related developments in the literature and the function is being widely studied. Some of these results are of immediate utility in forestry, and it appears the function will be a basis for future formulations. For example, diameter distributions, rather than following a simple unimodal form, sometimes are a mixture of two component distributions whose elements were not classified during data collection. The specification and estimation of parameters for such cases, utilizing a mixture of Weibull distributions, have been considered by several workers (Falls 1966, Kao 1959, Mason 1967). Censored data are also common in forestry, particularly Type I censoring on the left when trees below some threshold are counted but not accurately measured. Cohen (1965) gives background on censored samples from the Weibull distribution that is helpful in constructing a reasonable description of the distribution of all trees. For further discussion and a list of references the interested reader should consult Johnson and Kotz (1970). Elaboration concerning order statistics and the Weibull distribution is given by David (1970).

Estimating Parameters

Appropriate techniques for fitting the Weibull depend on statistical efficiency desired and available computing capabilities. Since many methods have been described, the selection of one for a given application can be a chore. A cross section of the various approaches to estimation is presented here, along with detailed background on two sets of estimators. One is based on maximum likelihood and the other is presented for simplicity of computations. We will assume in this discussion that the diameter data represent independent, random observations so as to utilize estimation theory. Other situations, such as when trees are selected with unequal probability or when the entire population is known and the problem is to summarize rather than estimate, are recognized but will not be discussed.

Maximum likelihood estimators are generally considered best, but for the Weibull the method of maximum likelihood requires iterative computations. Therefore, considerable emphasis has been placed on derivation of simpler estimators with acceptable properties.

Several methods have been developed, using transformations of the Weibull variable X to form a new variable $U = u(x;\ c, b)$, such that the expectations of order statistics $U_{(i)}$ from the distribution of U are independent of c and b. Estimators are then derived that minimize the squared deviations between transforms of observed and expected $U_{(i)}$. Variables must be transformed using Naperian logarithms. Computations then proceed much the same as for simple linear regression. Bain and Antle (1967) compared the estimators they derived, using

$$u(x;c,b) = (x/b)^c$$

with those of Gumbel (1958), Miller and Freund (1965), and Menon (1963). They concluded that Menon's estimators are better for complete data from large samples but unusable on censored data.

Other researchers have concentrated on obtaining linear functions to estimate c and $\ln(b)$ for complete or censored data. White (1969) gives weighted least squares estimators that are unbiased. Mann (1967) gives best linear estimators with the same large sample properties as maximum likelihood estimators. Both methods are applicable to censored data. However, they require tables of weights which must be entered for each sample size, degree of censoring, and ranked position.

D'Agostino (1971) gives a modification of a procedure derived by Johns and Lieberman (1966). The estimators are linear functions which are asymptotically jointly normal and efficient. The weights required are functions of the proportion of available observations and the sample size. Thus, for a complete sample the weights are simple functions of sample size. In a small sample comparison, D'Agostino shows that his estimators compare quite well with Mann's (1967) and White's (1969) estimators. They are more convenient because of the simplicity of the weight functions.

The easiest estimators to compute are based on percentiles. We define the 100-times-p^{th} percentile under the distribution as that value x_p of x such that a randomly chosen observation has probability p of being less than or equal to x_p. From equation (4),

$$\begin{aligned} p &= F(x_p) \\ &= 1 - \exp\{-x_p/b)^c\}. \end{aligned}$$

Thus,

$$x_p = b\,\{-\ln(1-p)\}^{1/c}. \tag{5}$$

With two sample percentiles x_r and x_t ($0 < r < 1;\ 0 < t < 1$), closed-form estimators for b and c can be obtained:

$$\begin{aligned} \hat{b} = \exp\{\,&[\ln(x_r) - \ln(x_t)\ln(-\ln(1-r)) \\ &/\ln(-\ln(1-t))] \\ &/\,[1 - \ln(-\ln(1-r)) \\ &/\ln(-\ln(1-t))\,]\} \end{aligned} \tag{6}$$

$$\hat{c} = \ln\{\ln(1-r)/\ln(1-t)\}\,/\ln\{x_r/x_t\}. \tag{7}$$

Dubey (1967) shows that the 17th and 97th sample percentiles are asymptotically best for estimating c with no prior knowledge of b, giving 66 percent efficiency when compared to the maximum likelihood estimator. The 40th and 82nd percentiles are best for estimating b for unknown c with 82 percent efficiency. He also shows that $\hat{b}$ and $\hat{c}$ using these percentiles are asymptotically normally distributed with smaller variances than Menon's (1963) estimators.

With c known, the 80th sample percentile is best and yields an asymptotically 65 percent efficient estimator for b. That estimator is

$$\hat{b} = \exp\{\ln(x_{.80}) - \ln(-\ln(.20))/c\}. \quad (8)$$

In such a case, however, the maximum likelihood estimator for b is easily obtained and has better properties.

Estimation of Weibull parameters by maximum likelihood has been covered in detail by Cohen (1965). He gives the likelihood equation, estimators, and variances and covariances of estimates for complete, singly censored, and progressively censored samples. With no censoring, obtaining the estimates for a given sample $(x_1, x_2, \ldots x_n)$ first requires the solution of

$$\{\sum_{i=1}^{n} x_i^{c^*} \ln(x_i)\}/\{\sum_{i=1}^{n} x_i^{c^*}\} - (1/c^*) = (1/n)\{\sum_{i=1}^{n} \ln(x_i)\} \quad (9)$$

for c^*, the estimator for c. Then

$$b^* = \{(1/n) \sum_{i=1}^{n} x_i^{c^*}\}^{1/c^*} \quad (10)$$

gives the estimator for b.

Attempting to solve equation (9) for c^* by the usual iterative processes can be time consuming if done by hand computation. However, Harter and Moore (1965) give a computerized iterative procedure for obtaining the maximum likelihood estimates of a, b, and c in model (2) using the first m order statistics from a sample of size n $(m \leqslant n)$. The method also allows any parameter or any pair of parameters to be estimated with the others specified in advance. A computer program in FORTRAN was obtained from Harter and modified to estimate Weibull parameters for examples in this paper.

Mann (1967) notes that c^* and b^* are consistent and asymptotically efficient, unbiased, and normally distributed. Thus, for large samples they are particularly good estimators. Even for small samples they are consistent, and correction for bias is possible. Thoman *et al.* (1969) give equa-

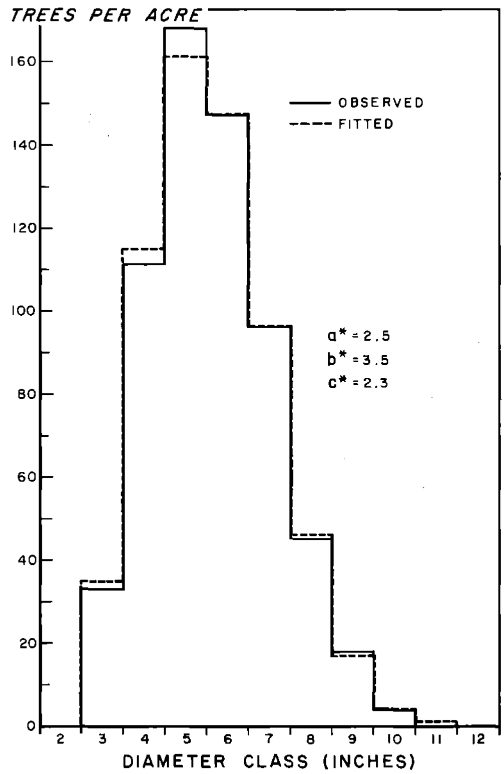

FIGURE 2. *Observed diameter distribution in 55-year-old jack pine stand and distribution fitted by three-parameter Weibull model.*

tions, graphs, and tables which make it possible to:

(1) Place exact confidence intervals on the parameters.

(2) Correct for bias.

(3) Test hypotheses regarding b and c and calculate the power of the test regarding c.

(4) Determine sample sizes at which asymptotic theory may apply.

Our investigation of presently available estimation methods has led to the conclusion that one should use the percentile estimators derived by Dubey (1967) when restricted to hand computations. Using equations (6) and (7) and appropriate values for r and t, estimates can be easily calculated when tables of logarithms are the only computational aids available. With access to an adequate computer, maximum likelihood estimation is the best choice.

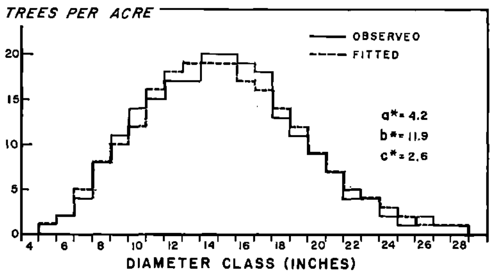

FIGURE 3. *Observed diameter distribution in 67-year-old Douglas-fir stand and distribution fitted by three-parameter Weibull model.*

Harter and Moore's (1965) iterative procedure is very efficient, relatively easy to program, and designed to handle cases where the usual iterative processes break down. Some of the other procedures, perhaps the one by D'Agostino (1971), will find utility in the middle ground of computational capabilities.

Examples

As examples, four published diameter distributions were fitted with the Weibull function. Two mound-shaped distributions (Figs. 2 and 3) were fitted by equation (2) and maximum likelihood.

The first example (Fig. 2) represents an even-aged stand of jack pine at age 55 in the Cloquet Experimental Forest, Cloquet, Minnesota (Spurr 1952, p. 355). The maximum difference between observed and fitted distributions is 7 trees and occurs in the 5-inch class. This difference represents only 4 percent of the observed trees per acre for that class.

Data from a 67-year-old Douglas-fir stand, site quality III (Chapman and Meyer 1949, p. 14) were fitted even closer (Fig. 3). The maximum difference of two trees per acre between observed and fitted frequencies occurs in diameter classes 10, 13, 16, and 17. The expected minimum diameter is 4.2 inches, and 63 percent of the stems in the fitted distribution are less than 16.1 inches.

Examples of reversed J-shaped distributions were found in Davis (1966, p 216). The data are from a stand in the shortleaf-loblolly pine-hardwood type on the Crossett Experimental Forest in Arkansas. The stand is shown first (Fig. 4a) in a severely understocked condition brought about by fire and heavy cutting. It is then presented (Fig. 4b) after 10 years of management. The numbers of stems by diameter class were so low in

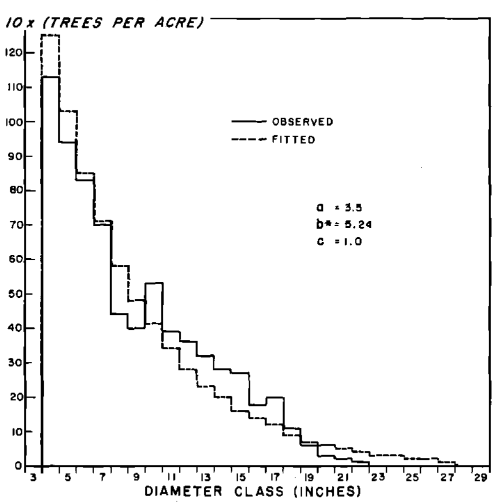

FIGURE 4a. *Observed and fitted J-shaped diameter distributions. Understocked shortleaf-loblolly pine-hardwood stand.*

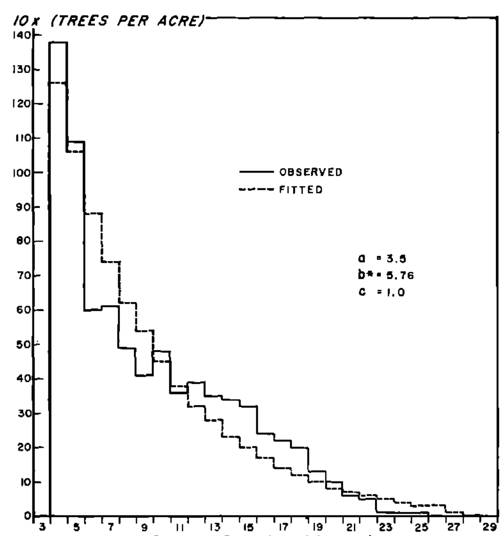

FIGURE 4b. *Same stand as in Fig 4a after 10 years of management.*

the actual data—the class with the highest frequency had 13.8 stems per acre—that each number was multiplied by 10 before estimating the Weibull parameters.

The unrestricted fit by maximum likelihood declared these two distributions to be extremely right-skewed mounds rather than reversed J-shaped because there was no sharp drop in number of stems between the first two classes. The data were refitted with c^* specified in advance to be 1.0, forcing the maximum likelihood estimation procedure to fit an exponential curve. If one were calculating the estimates by percentiles, equation (8) would be used with $c = 1.0$. The smallest observed diameter could be used as an estimate of a.

In a dissertation[1] results are presented which indicate good fits to data from even-aged stands and well-defined relationships between stand characteristics and percentiles of the fitted Weibull distributions. Parameters of (2) were estimated by maximum likelihood with remeasurement data under the constraint that parameter a not decrease with increasing stand age.

Conclusions

Many models for diameter distributions have been proposed, but none exhibit as many desirable features as the Weibull. The divers parameter estimation procedures for this function allow a selection appropriate to computing capabilities. Simplicity of algebraic manipulations and ability to assume a variety of curve shapes should make the Weibull useful for other biological models.

Literature Cited

BAIN, L. J., and C. E. ANTLE. 1967. Estimation of parameters in the Weibull distribution. Technometrics 9:621–627.

BLISS, C. I., and K. A. REINKER. 1964. A lognormal approach to diameter distributions in even-aged stands. Forest Sci 10:350–360.

CHAPMAN, H. H., and W. H. MEYER. 1949. Forest mensuration. 522 p. McGraw-Hill, N.Y.

CLUTTER, J. L., and F. A. BENNETT. 1965 Diameter distributions in old-field slash pine plantations. Ga For Res Counc Rep 13, 9 p

COHEN, A. C., JR. 1965. Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. Technometrics 7:579–588.

D'AGOSTINO, R. B. 1971. Linear estimation of the Weibull parameters. Technometrics 13· 171–182.

DAVID, H. A. 1970. Order statistics. 272 p John Wiley & Sons, N.Y.

DAVIS, K. P. 1966. Forest management: regulation and valuation. 519 p. Second ed, McGraw-Hill, N.Y.

DUBEY, S. D. 1967. Some percentile estimators for Weibull parameters. Technometrics 9:119–129.

FALLS, L. W. 1966. Estimation of parameters in compound Weibull distributions. NASA Tech Memo TMX-53422.

FISHER, R. A., and L. H. C. TIPPETT. 1928 Limiting forms of the frequency distribution of the largest or smallest member of a sample *In* 24th Camb Philos Soc Proc 1927:180–190.

GUMBEL, E. J. 1958. Statistics of extremes 375 p. Columbia University Press, N.Y.

HARTER, H. L. 1964. Expected values of exponential, Weibull and gamma order statistics. Off Aerosp Res, US Air Force, Proj 7071, Task 7071-01, ARL 64-31, 106 p.

———, and A. H. MOORE. 1965. Maximum likelihood estimation of the parameters of gamma and Weibull populations from complete and from censored samples. Technometrics 7:639–643.

JOHNS, M. V., JR., and G. J. LIEBERMAN 1966. An exact asymptotically efficient confidence bound for reliability in the case of the Weibull distribution. Technometrics 8 135–175.

JOHNSON, NORMAN L., and SAMUEL KOTZ. 1970 Continuous univariate distributions—1. 300 p Houghton Mifflin Co., Boston.

KAO, J. H. K. 1959. A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. Technometrics 1:389–407.

LEAK, W. B. 1965. The J-shaped probability distribution. Forest Sci 11:405–409.

LENHART, J. DAVID, and J. L. CLUTTER. 1971 Cubic-foot yield tables for old-field loblolly pine plantations in the Georgia Piedmont. Ga For Res Counc Rep 22, series 3, 13 p.

MANN, N. R. 1967. Tables for obtaining the best linear invariant estimates of parameters of

[1] Bailey, R. L. Development of unthinned stands of *Pinus radiata* in New Zealand. 67 p. Univ. Ga. Ph.D. Diss. 1972.

the Weibull distribution. Technometrics 9: 629–645.

———. 1968. Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions. Technometrics 10: 231–256.

MASON, T. J. 1967. Maximum likelihood estimation in a mixture of two Weibull distributions. Univ Ga Inst Stat Tech Rep 19, 25 p.

MCGEE, C. E., and L. DELLA-BIANCA. 1967. Diameter distributions in natural yellow-poplar stands. USDA Forest Serv Res Pap SE-25, 7 p. Southeast Forest Exp Stn, Asheville, NC.

MENON, M. V. 1963. Estimation of the shape and scale parameters of the Weibull distribution. Technometrics 5:175–182.

MEYER, H. A. 1952. Structure, growth, and drain in balanced uneven-aged forests. J Forest 50:85–92.

———, and D. D. STEVENSON. 1943. The structure and growth of virgin beech-birch-maple-hemlock forests in northern Pennsylvania. J Agric Res 67:465–484.

MEYER, W. H. 1930. Diameter distribution series in evenaged forest stands. Yale Univ Sch For Bull 28, 105 p.

MILLER, I., and J. E. FREUND. 1965. Probability and statistics for engineers. 432 p. Prentice-Hall, Englewood Cliffs, N.J.

NELSON, T. C. 1964. Diameter distribution and growth of loblolly pine. Forest Sci 10: 105–115.

OSBORNE, J. G., and F. X. SCHUMACHER. 1935. The construction of normal-yield and stand tables for even-aged timber stands. J Agric Res 51:547–564.

POLLARD, J. H. 1971. On distance estimators of density in randomly distributed forests. Biometrics 27:991–1002.

SCHMELZ, D. V., and A. A. LINDSEY. 1965 Size-class structure of old-growth forests in Indiana. Forest Sci 11:258–264.

SCHNUR, G. L. 1934. Diameter distributions for old-field loblolly pine stands in Maryland J Agric Res 49:731–743.

SPURR, S. H. 1952. Forest inventory. 476 p Ronald Press, N.Y.

THOMAN, D. R., L. J. BAIN, and C. E. ANTLE 1969. Inferences on the parameters of the Weibull distribution. Technometrics 11:445–460.

WEIBULL, W. 1939. A statistical theory of the strength of material. Ing Vetensk Akad Handl 153, 17 p.

———. 1951. A statistical distribution function of wide applicability. J Appl Mech 18: 293–297.

WHITE, J. S. 1969. The moments of log-Weibull order statistics. Technometrics 11: 373–386.

Erratum

IN THE NOTE by L. P. Abrahamson and L. Newsome, "Tree Age Influences Trunk Borer Infestations in Cottonwood Plantations," *Forest Science* 18:231–232, the plotted line in Figure 1 can be accurately read in terms of percent of trees infested. In the caption to the Figure, however, the second sentence should be deleted and replaced by: $Y = \text{arcsine}\ (\sqrt{\text{Proportion of trees infested}})$, $X = \text{age}$, and 1/age was used as the weight in fitting the regression.