United States
Department of
Agriculture

Forest Service

Southern
Research Station

Research Paper
SRS–31

Recovering Parameters of Johnson's S_B Distribution

Bernard R. Parresol

The Author

Bernard R. Parresol, Mathematical Statistician, Southern Forest Inventory, Monitoring, and Analysis Program, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC 28804.

April 2003

Southern Research Station
P.O.Box 2680
Asheville, NC 28802

Recovering Parameters of Johnson's S_B Distribution

Bernard R. Parresol

Abstract

A new parameter recovery model for Johnson's S_B distribution is developed. This latest alternative approach permits recovery of the range and both shape parameters. Previous models recovered only the two shape parameters. Also, a simple procedure for estimating the distribution minimum from sample values is presented. The new methodology employs the median and the first and second moments of the distribution. The methodology is demonstrated by modeling diameter distributions of unthinned loblolly pine plantations. Compatible equations for projecting per-hectare values of number of trees and basal area from initial stand conditions are presented, as well as equations for predicting median diameter, mean diameter, and the location parameter. Given estimates of these five stand attributes, the range and the two shape parameters of the S_B distribution can be recovered. The χ^2 goodness-of-fit test rejected 56 cases out of 527 for conformance to an S_B distribution. Though the S_B distribution is very flexible in terms of distribution shape, about 10 percent of the loblolly plantation observations did not follow this distribution. Nonetheless, deviation analysis showed reasonable results, with 77 percent of the variation explained in current and projected distributions (numbers of trees by 2.5-cm diameter class). Overall, the recovered S_B distributions provided good approximations of the observed diameter distributions.

Keywords: Diameter distributions, kurtosis, Newton-Raphson method, moments, parameter recovery model, percentiles, *Pinus taeda*, skewness.

Introduction

Forest managers have long relied on growth-and-yield forecasts in making management decisions. Growth-and-yield models that provide detailed stand distributional information are particularly useful. Clutter and Bennett (1965) introduced diameter distribution methodology using the four-parameter beta probability density function (PDF) to describe the distribution of the number of trees per unit area by diameter at breast height (d.b.h.) class. Gove and Patil (1998) refer to this as the d.b.h.-frequency distribution. Bailey and Dell (1973) were the first to use the Weibull probability density function as a d.b.h.-frequency distribution model. They justified use of the Weibull over the beta distribution because the Weibull has a closed form expression for its cumulative distribution function (CDF) and only three parameters. Hafley and Schreuder (1977) introduced the S_B (system bounded) distribution (Johnson 1949) to the forestry literature.

The original approach for developing a diameter distribution model involved (1) estimating the PDF parameters from sample data using either the method of moments or maximum likelihood and (2) development of regression equations to relate the parameter estimates to the stand variables age, site, and density. These so-called parameter prediction models (Hyink 1980, Hyink and Moser 1983) were straightforward to fit. Unfortunately, functions for relating the PDF parameters to stand characteristics usually accounted for only a small percentage of the variation in the parameter estimates (i.e., low R^2 values; see Feduccia and others 1979, Smalley and Bailey 1974). The alternative put forth by Hyink (1980) was the parameter recovery method. In this method, the parameters of the distribution function are solved (recovered) from a system of equations derived by equating measured or predicted forest stand attributes to their analytical counterparts from the PDF. Parameter recovery models are most often based on percentiles of diameter distribution or moments of diameter distribution. The parameter recovery method, in general, is superior to the parameter prediction method for the projection of future distribution parameters because d.b.h.-frequency distribution characteristics, such as mean diameter and diameter variance, can be projected with more confidence than the distribution parameters themselves. In many growth-and-yield studies the two methods—parameter prediction and parameter recovery—are used simultaneously. That is, some of the PDF parameters are predicted, and others are solved using a parameter recovery approach. Over the past two decades there have been many diameter distribution growth-and-yield systems based on parameter recovery (possibly in combination with parameter prediction) using the Weibull distribution (e.g., Baldwin and Feduccia 1987, Cao and others 1982, Gove and Patil 1998, Matney and Sullivan 1982, Zarnoch and others 1991), and to a lesser extent the S_B distribution (e.g., Hafley and Buford 1985, Scolforo and Thierschi 1998, Tham 1988). The goal of this article is to describe the development and testing of a new alternative approach for parameter recovery in the S_B distribution.

The S_B Distribution

Johnson (1949) realized that a random variable X that is bounded (i.e., has upper and lower limits) could be transformed to approximate normality by the transformation

$$z = \gamma + \delta \ln\left[(x-\xi)/(\xi+\lambda-x)\right] = g^{-1}(x)$$

where ln is the natural logarithm and $Z \sim \mathcal{N}(0,1)$. Now Z has PDF

$$n(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$$

where exp is the base of the natural logarithm. The equation for Johnson's S_B distribution therefore is

$$f(x) = n(g^{-1}(x))\,|\,dg^{-1}(x)/dx\,|$$

or

$$f(x) = \begin{cases} \dfrac{\delta}{\sqrt{2\pi}} \dfrac{\lambda}{(x-\xi)(\xi+\lambda-x)} \exp\left\{ -\dfrac{1}{2}\left[\gamma + \delta \ln\left(\dfrac{x-\xi}{\xi+\lambda-x} \right) \right]^2 \right\}, \\ \xi < x < \xi+\lambda, \delta > 0, -\infty < \gamma < \infty, \lambda > 0, \xi \geq 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{1}$$

It is characterized by the location parameter ξ, the range parameter λ, and shape parameters γ and δ.

Hafley and Schreuder (1977) examined the skewness coefficient, $\sqrt{\beta_1}$, and kurtosis coefficient, β_2, of various statistical distributions as a measure of the flexibility of the distributions in regard to their changes in shape. Skewness is defined as

$$\sqrt{\beta_1} = \mu_3 / \mu_2^{3/2}$$

and kurtosis as

$$\beta_2 = \mu_4 / \mu_2^2$$

where

$$\mu_r = \int_{-\infty}^{\infty} \left[x - E(X) \right]^r f(x)dx \text{ and } f(x) \text{ is the PDF of } X$$

Skewness is a departure from symmetry about the mean, where negative values indicate a distribution with a longer tail to the left and positive values a longer tail to the right of the mean. Kurtosis is a measure of the heaviness of the tails of a distribution; the larger the value of β_2 the more sizeable the tails. There is a myth that kurtosis measures the *peakedness* of a density, despite repeated examples to the contrary (Ali 1974, Johnson and others 1980, Kaplansky 1945). For the standard normal, $Z \sim \mathcal{N}(0,1)$, $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$. Moment estimators of $\sqrt{\beta_1}$ and β_2 are

$$\sqrt{b_1} = \frac{\sum (x_i - \bar{x})^3}{\left[\sum (x_i - \bar{x})^2 \right]^{3/2}} \cdot \sqrt{n} \tag{2}$$

$$b_2 = \frac{\sum (x_i - \bar{x})^4}{\left[\sum (x_i - \bar{x})^2 \right]^2} \cdot n \tag{3}$$

where n is the sample size.

In figure 1, a graph of the β_1–β_2 space is presented showing the Weibull, gamma, and lognormal distributions. By convention the ordinate (β_2) scale is presented upside down. Certain combinations of β_1 and β_2 are mathematically impossible and occur in the region above the line $\beta_2 - \beta_1 - 1 = 0$. The three distributions shown are represented by lines in the β_1–β_2 space. Because the graph of figure 1 presents β_1, the square of the skewness coefficient, the positive and negative aspect of a distribution is not obvious. The lower line of the Weibull plot in figure 1 is generated by negatively skewed shapes. The beta distribution covers the region between the gamma line, the impossible region, and the β_2 axis. Johnson's S_B distribution covers the region between the lognormal line, the impossible region, and the β_2 axis. Hence, the beta and S_B distributions cover a broad spectrum of shapes, fitting both positively and negatively skewed data.

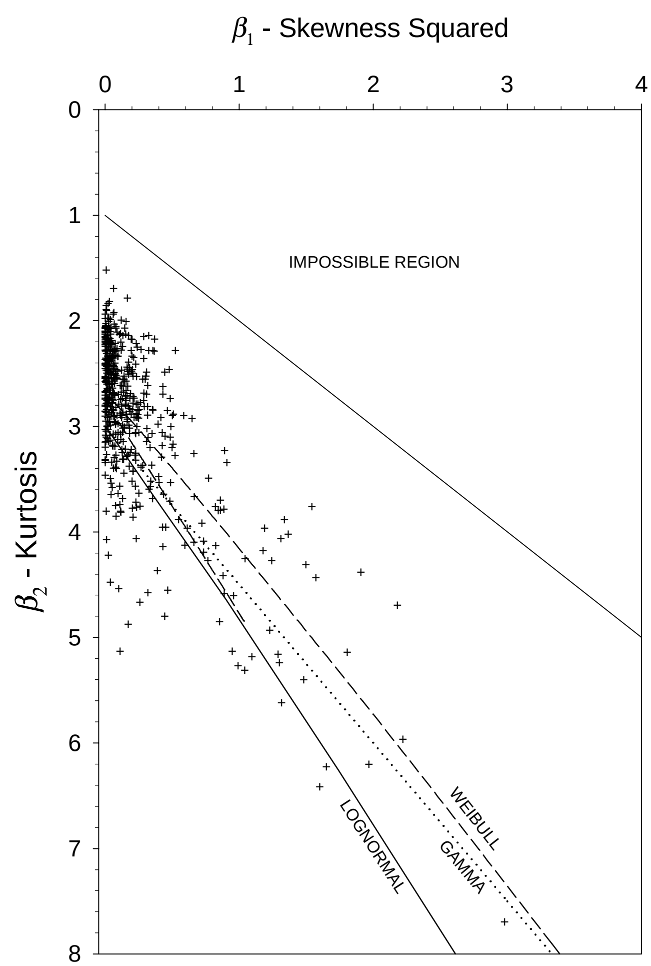

Figure 1—The β_1–β_2 space. The beta distribution covers the region between the gamma line and the impossible region. The S_B distribution covers the region between the lognormal line and the impossible region. The + symbols mark skewness squared-kurtosis values for the 527 unthinned loblolly pine plot observations.

Because the S_B distribution is obtained by a transformation on a standard normal variate, integration over specific classes can be accomplished by application of the well-tabulated standard normal. Further, the distribution can easily be extended to multivariate forms (Hafley and Buford 1985, Knoebel and Burkhart 1991). These factors plus the distribution's flexibility in the β_1–β_2 space were the salient points Hafley and Schreuder (1977) used to justify selection of the S_B distribution over other distributions.

Previous S_B Models

In previous S_B-based models the general approach has been to specify the distribution minimum and maximum values, thereby fixing $\hat{\xi}$ = minimum value and $\hat{\lambda}$ = maximum value – minimum value, and to recover or otherwise estimate the two shape parameters. Hafley and Buford (1985) presented a modeling approach whereby they used a maximum likelihood type estimating equation for δ, that is,

$$\hat{\delta} = \frac{\hat{\lambda}}{4s_x}$$

where s_x is the standard deviation of x (which is a function of the first and second moments of a distribution), and for γ they used

$$\hat{\gamma} = \frac{2x_m - \hat{\xi} - \hat{\lambda}}{\hat{\lambda}\hat{\delta}} - \hat{\delta}\ln\left(\frac{x_m - \hat{\xi}}{\hat{\lambda} + \hat{\xi} - x_m}\right)$$

where x_m is the mode of x. Other researchers have relied on the two-percentile method to recover γ and δ. In Newberry and Burk (1985) and Knoebel and Burkhart (1991) the 50th and 95th percentiles were used. From the formulation of the S_B distribution,

$$Z_{95} = \gamma + \delta\ln\left(\frac{X_{95} - \xi}{\xi + \lambda - X_{95}}\right) \text{ and } Z_{50} = \gamma + \delta\ln\left(\frac{X_{50} - \xi}{\xi + \lambda - X_{50}}\right)$$

where Z_{95} and Z_{50} represent the standard normal values corresponding to the cumulative frequencies of 95 percent and 50 percent of the standard normal distribution, and X_{95} and X_{50} are the 95th and 50th percentiles of the observed distribution. Because Z_{50} = 0, we have

$$\hat{\gamma} = -\hat{\delta}\ln\left(\frac{X_{50} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - X_{50}}\right)$$

Substituting into the equation for Z_{95} and simplifying yields

$$\hat{\delta} = \frac{1.645}{\ln\left(\frac{X_{95} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - X_{95}}\right) - \ln\left(\frac{X_{50} - \hat{\xi}}{\hat{\xi} + \hat{\lambda} - X_{50}}\right)}$$

Newbury and Burk (1985) correctly pointed out that these two percentile points, the 50th and 95th, do not necessarily provide the optimal estimates for γ and δ.

Of course, many approaches have been used to determine the S_B parameters, such as the four-percentile points method, linear and nonlinear regression methods, and maximum likelihood. These have been reviewed and compared by Zhou and McTague (1996) and Kamziah and others. (1999). The purpose of my brief review of the above methodology is to provide a contrast against the methodology presented in the next section.

Development of the S_B Parameter Recovery System

The parameter recovery method ensures compatibility between stand characteristics generated from a distribution function and predicted from regressions. To apply the parameter recovery method, a system of equations involving certain tree stand characteristics and the S_B parameters must be developed. If we define a new variable Y with variate values

$$y = (x - \xi)/\lambda \tag{4}$$

then the S_B distribution [equation (1)] can be expressed as

$$f(y) = \frac{\delta}{\sqrt{2\pi}}\frac{1}{y(1-y)}\exp\left\{-0.5\left[\gamma + \delta\ln\left(\frac{y}{1-y}\right)\right]^2\right\},\ 0 < y < 1$$

where

$$z_y = \gamma + \delta\ln[y/(1-y)] \tag{5}$$

Solving equation (5) for y gives

$$y = (1 + e^{-(z-\gamma)/\delta})^{-1} = g(z) \tag{6}$$

Setting z = 0 in equation (6) results in the median value of Y:

$$y_{\text{median}} = (1 + e^{\gamma/\delta})^{-1}$$

hence

$$\gamma = \delta\ln(1/y_{\text{median}} - 1) \tag{7}$$

Because Y is a function of Z [see equation (6)], $E(Y^r) = E[g(Z)^r]$. As Z is a unit normal variable, the rth noncentral moment of Y is

$$\mu_r'(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-z^2/2)/[1 + \exp((\gamma - z)/\delta)]^r\,dz \tag{8}$$

which can be evaluated using a formula due to Goodwin (1949). Hill and others (1976) give an excellent algorithm for evaluating S_B moments based on Draper's (1952) form of Goodwin's integral.

Let us proceed by solving equation (4) for *x*, which gives $x = \lambda y + \xi$. Taking the statistical expectation [i.e., $E(X) = E(\lambda Y + \xi)$] we obtain

$$\bar{x} = \lambda \mu_1'(y) + \xi \tag{9}$$

Quadratic mean d.b.h. ($\bar{D}_q$), number of trees per unit area (*N*), and basal area per unit area (*B*) are related by the equation (see Clutter and others 1983, p. 89) $\bar{D}_q = \sqrt{B/(\mathcal{K} N)}$, thus $B = \mathcal{K} N \cdot \bar{D}_q^2$ where $\mathcal{K}$ is for units conversion from d.b.h.^{2} to basal area ($\mathcal{K} = \pi / 40{,}000$ for converting d.b.h.^{2} in cm^2 to area in m^2) and $\bar{D}_q^2 = E(X^2)$. In terms of our new variable *Y*, $E(X^2) = E[(\lambda Y + \xi)^2] = \lambda^2 \mu_2'(y) + 2\xi\lambda\mu_1'(y) + \xi^2$. Now *B*, expressed as a function of the first and second moments of *Y*, becomes

$$B = \mathcal{K} N [\lambda^2 \mu_2'(y) + 2\xi\lambda\mu_1'(y) + \xi^2] \tag{10}$$

The relationship in equation (7) is used to eliminate γ in equations (9) and (10) by substitution into equation (8). We are then left with a system of two nonlinear equations and two unknowns. Given estimates of *B*, *N*, median tree diameter (d_{median}), average tree diameter (d_{mean}), and the ξ parameter, plus the relationships in equations (9) and (10), we can iteratively solve for parameters λ and δ. The parameter γ is then determined from equation (7).

An attempt was made to define three equations in three unknowns so that the ξ parameter could be recovered as well. Taking the variance of equation (4) gives

$$\text{var}(X) = \lambda^2 \left[\mu_2'(y) - \mu_1'(y)^2 \right]$$

However, this function is a combination of equations (9) and (10), which is readily seen by subtracting the square of equation (9) from equation (10). A system based on these three equations cannot be solved. Next kurtosis was tried, which, expressed in terms of noncentral moments, is

$$\beta_2 = \frac{\mu_4'(y) - 4\mu_1'(y)\mu_3'(y) + 6\mu_1'(y)^2 \mu_2'(y) - 3\mu_1'(y)^4}{\left[\mu_2'(y) - \mu_1'(y)^2 \right]^2}$$

However, the resulting complexity of this system of three nonlinear equations posed such convergence problems that it was abandoned.

Undoubtedly overall stand attributes such as volume and basal area are less sensitive to errors in ξ than to errors in the other three parameters. Hence, specifying ξ and recovering the remaining parameters is probably the most expedient approach.

Solution Technique

The solution of a system of nonlinear equations is a difficult problem. A two-dimensional Newton-Raphson procedure is relatively straightforward to program and has quadratic convergence. The two-dimensional Newton-Raphson method can be written in vector notation as (Ralston and Rabinowitz 1978, p. 360)

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left(\frac{1}{g} \begin{bmatrix} \dfrac{\partial f_2}{\partial x^{(2)}} & -\dfrac{\partial f_1}{\partial x^{(2)}} \\ -\dfrac{\partial f_2}{\partial x^{(1)}} & \dfrac{\partial f_1}{\partial x^{(1)}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \right)_{\mathbf{x}=\mathbf{x}_i} \tag{11}$$

where the determinant *g* is

$$g = \begin{vmatrix} \dfrac{\partial f_1}{\partial x^{(1)}} & \dfrac{\partial f_1}{\partial x^{(2)}} \\ \dfrac{\partial f_2}{\partial x^{(1)}} & \dfrac{\partial f_2}{\partial x^{(2)}} \end{vmatrix}$$

The key to using equation (11) is correctly specifying the partial derivatives of f_1 and f_2 with respect to $x^{(1)}$ (the λ parameter) and $x^{(2)}$ (the δ parameter). I shall denote the relationship in equation (9) as f_1 and in equation (10) as f_2. The partial derivatives of equations (9) and (10) depend on the partial derivatives of the moments of *Y*. Let us define the pseudo-moment $\tilde{\mu}_r'(y)$ as

$$1/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp(-z^2/2) / [1 + \exp((\gamma - z)/\delta)]^r \, z \;\; dz \tag{12}$$

The partial derivatives of the moments of *Y* (see appendix) can be shown to be functions of $\mu_r'(y)$ and $\tilde{\mu}_r'(y)$. Using the results of equations (A1) and (A2), the partial derivatives of f_1 and f_2 are:

$$\frac{\partial f_1}{\partial \lambda} = \mu_1'(y) - \frac{\lambda\left(\mu_1'(y) - \mu_2'(y)\right)}{\lambda - d_{\text{median}} + \xi} \tag{13}$$

$$\frac{\partial f_1}{\partial \delta} = \frac{-\lambda\left(\tilde{\mu}_1'(y) - \tilde{\mu}_2'(y)\right)}{\delta^2} \tag{14}$$

$$\frac{\partial f_2}{\partial \lambda} = \left\{ \frac{-2\lambda\left[\lambda\left(\mu_2'(y)-\mu_3'(y)\right)+\xi\left(\mu_1'(y)-\mu_2'(y)\right)\right]}{\lambda - d_{\text{median}} + \xi} + 2\left(\lambda\mu_2'(y)+\xi\mu_1'(y)\right) \right\} \cdot \mathcal{K}\, N \tag{15}$$

$$\frac{\partial f_2}{\partial \delta} = \left\{ \frac{-2\lambda\left[\lambda\left(\tilde{\mu}_2'(y)-\tilde{\mu}_3'(y)\right)+\xi\left(\tilde{\mu}_1'(y)-\tilde{\mu}_2'(y)\right)\right]}{\delta^2} \right\} \cdot \mathcal{K}\, N \tag{16}$$

Substituting equations (13) through (16) into equation (11) gives us the two-dimensional Newton-Raphson iteration formula to solve the system of equations. The L_1 norm

$$\| \mathbf{f} \|_1 = \sum_{i=1}^{2} | f_i |$$

was used to define convergence as

$$\| \mathbf{f} \|_1 < 0.002$$

Data

Data came from 287 plots (varying in size from 0.04 to 0.3 ha) established in unthinned loblolly pine (*Pinus taeda* L.) plantations in north and central Louisiana, southern Arkansas, southeast Texas, and southwest Mississippi. Periodic remeasurements of stand conditions were taken every 4 to 6 years on 125 of the plots. The database consisted of 527 stand observations, 240 of which were growth observations (remeasurements). Diameter at breast height (1.37 m) to the nearest 0.25 cm was measured for each tree on the plot. The average height of dominant and codominant trees was obtained on each plot at each measurement to determine site index. Stand ages ranged from 5 to 45 years, trees per hectare ranged from 119 to 3,039, basal area per hectare varied from 1.15 to 48.2 m^2, and average height of dominants and codominants varied from 5.2 to 27.4 m.

Figure 1 presents a plot of the 527 (b_1, b_2) values, calculated from equations (2) and (3) on the diameter measurement values. Few points fall near the line defined by the Weibull distribution. Clearly, the S_B distribution covers most of the (b_1, b_2) values and therefore should be appropriate for modeling these stand diameter distributions.

The 0th Percentile for the Distribution Minimum Value

Estimates of the location parameters of the recovered distributions are needed. In previous studies (Cao and others 1982, Matney and Sullivan 1982, Zöhrer 1972) regressions were developed on the sample plot minimum diameters, then subjective adjustments, such as dividing the prediction by 2, were made to estimate the threshold parameters. Other studies (Dell and others 1979, Feduccia and others 1979, Smalley and Bailey 1974) developed direct estimation of threshold parameters from regressions on maximum likelihood estimates (MLEs). Objective methods are preferable, so maximum likelihood was used to solve for the 527 ξ parameters. However, the MLEs were unsatisfactory. Values of $\hat{\xi}$ were nearly 0 in most cases and often were in conflict with the natural progression of diameter over time.

Looking at plots of the empirical CDFs, I observed that the curves tended to straighten at the ends. I took the three smallest plot diameter values and linearly regressed down to the 0th percentile value (fig. 2). Percentiles were calculated on the ordered d.b.h. values d_1, d_2, . . ., d_n as $p(d_i) = i/(n+1) \times 100$. This technique gave reasonable diameter values for the minimum distribution values.

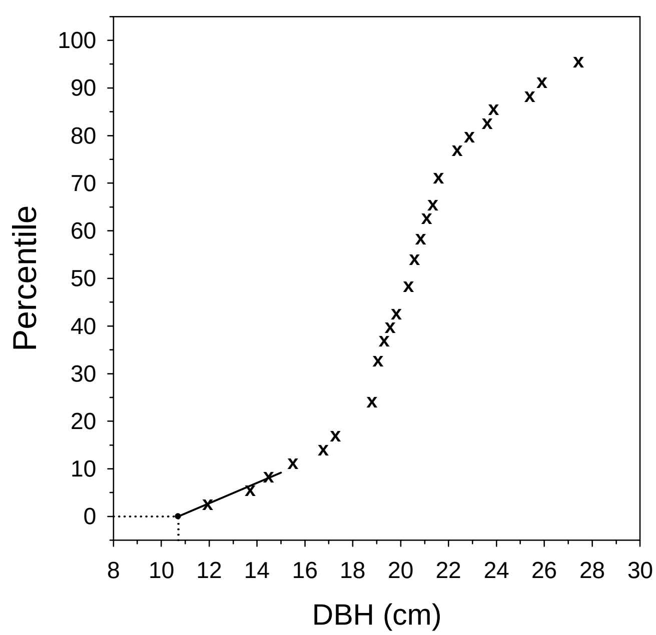

Figure 2—Empirical CDF of a loblolly pine study plot with a linear regression line fitted to the three smallest values showing extrapolation to the 0th percentile value.

Stand Equations

Compatible projection models for number of trees surviving and basal area were fitted to the sample data. Additionally, prediction equations for median d.b.h., mean d.b.h., and the ξ parameter also were fitted. Table 1 shows the equations that predict stand attributes. The loblolly pine site index curve used was that of Popham and others (1979).

In addition to stand attribute values, starting values are required to begin the iteration of equation (11). For the range parameter, the following equation gave good starting values:

$$\text{initial } \lambda = 12.29 + 0.2562 \cdot s_x^2 - 0.0123 \cdot (N/A) + 0.00338 \cdot s_x^2 \cdot N/A$$

where s_x^2 is the variance of d.b.h. in cm², N is number of trees per hectare, and A is stand age in years. For δ, the value 1.2 was determined to be a good starting point, based on the observation that δ often occurs between 1 and 1.5 for S_B diameter distributions.

Results

A FORTRAN program was written to facilitate use of the parameter recovery model and to provide a means of

Table 1—Equations for predicting unthinned loblolly plantation stand attributes

Equation[a]
$\hat{N}_2 = N_1 - 1.05409 \cdot 10^{-6}(A_2 - A_1)^{1.06412} \cdot (N_1/B_1)^{2.38353} \cdot HD_1^{2.47812} \cdot \exp(4.56524 \cdot 10^{-2} \cdot B_1)$
$n = 240;\ FI^{\text{b}} = 0.964;\ \text{RMSE} = 83.5$
$\hat{B}_2 = B_1 + 405.39237(A_2 - A_1) \cdot B_1^{0.42213} \cdot HD_1^{4.38106 \cdot 10^{-2}} \cdot A_2^{-2.49276} \cdot \exp(-5.10433 \cdot 10^{-21} \cdot B_1^{12.69457})$
$n = 240;\ FI = 0.942;\ \text{RMSE} = 2.3674$
$\hat{d}_{\text{median}} = -0.51862 + 1.02175\bar{D}_{\text{q}} - 1.60392 \cdot 10^{-2} \cdot A \cdot I \quad (I = 1 \text{ if } A \geq 20 \text{ else } I = 0)$
$n = 527;\ FI = 0.988;\ \text{RMSE} = 0.5428$
$\widehat{\ln(\bar{D}_{\text{q}} - d_{\text{mean}})} = -1.95405 + 0.43658 \ln(HD)$
$n = 527;\ FI = 0.999;\ \text{RMSE} = 0.1732$
$\hat{\xi} = -0.90155 + 0.018167\bar{D}_{\text{q}}^2 - 0.27733(\bar{D}_{\text{q}}^2 - d_{\text{mean}}^2) + 0.43026HD$
$n = 523;\ FI = 0.757;\ \text{RMSE} = 2.1996$

[a] Notation: A = stand age in years, B = basal area in m² per hectare, d_{mean} = average d.b.h. in cm, d_{median} = median d.b.h. in cm, $\bar{D}_{\text{q}}$ = quadratic mean d.b.h. in cm, HD = average height of dominants and codominants in m, N = number of surviving trees per hectare, and ξ = distribution minimum d.b.h. in cm. Subscripts refer to time of measurement.

[b] Statistics: fit index, $FI = 1 - \sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y}_i)^2$; root mean square error, RMSE = $\sqrt{\sum(y_i - \hat{y}_i)^2/(n-p)}$; where n = number of observations, p = number of parameters, y_i = observed value, $\bar{y}$ = observed mean value, and $\hat{y}_i$ = predicted value.

checking the model against the data. Inputs into the program were initial age, initial basal area, initial number of trees, initial height of dominant and codominant trees or base age 25 site index, the number of projection periods, and the time length between projection periods. The program provided the recovered parameter values, current and future stand tables, and histograms.

Initial stand conditions for each of the 287 plots were input into the program, and 287 current and 240 future stand tables were generated and compared against the observed distributions. A convergent solution was obtained on all 527 observations. On the average, convergence was obtained in 3 to 4 iterations. From a purely mechanical viewpoint, the S_B parameter recovery model worked extremely well. The degree of correspondence between the actual and predicted d.b.h.-frequency distributions was evaluated next.

For all observations, differences were calculated between observed and predicted number of trees by 2.5-cm d.b.h. classes. The generated S_B distributions accounted for 77 percent of the observed variation in number of trees by d.b.h. class. The standard deviation was 67 trees per hectare. Residual analysis showed no apparent prediction bias except in the smallest diameter class, where numbers were slightly underestimated.

The χ^2 goodness-of-fit test, using 2.5-cm classes, was performed on each of the observations. These tests resulted in rejection of the hypothesis that diameter distribution is from an S_B distribution in 56 cases at the $\alpha = 0.01$ level. While this number is considerably larger than the expected number of rejections $(0.01 \times 527 \cong 5)$, it was not unanticipated. The graph of skewness squared-kurtosis values in figure 1 clearly shows some values outside the range covered by the S_B distribution. Of the 56 cases rejected, 50 had values [i.e., (b_1, b_2) points] that fell below the lognormal line. Taking these 50 observed plot distributions into account $(56 - 50 = 6)$, the χ^2 test results are very reasonable. It was concluded, therefore, the recovered S_B distributions provided good approximations of the observed diameter distributions.

Discussion

A parameter recovery framework was developed using the stand level values of mean d.b.h. and basal area, in conjunction with median d.b.h., and the first and second moments of the S_B distribution. The resulting system of two nonlinear equations [equations (9) and (10)] plus equation (7), which form the parameter recovery model, allowed for the determination (i.e., recovery) of the range and both shape parameters of the S_B distribution. The approach of Hafley and Buford (1985) and the two-percentile method used by Newbury and Burk (1985) and Knoebel and Burkhart (1991) do not require iterative procedures. However, Kamziah and others (1999) were critical of the two-percentile method because ". . . the parameters γ and δ were estimated based only on two percentiles, and λ was not estimated simultaneously from these percentiles." They felt it was important to interconnect the value of the range parameter to the overall shape of the distribution, because the range and shape parameters provide feedback to each other (a gestalt, if you will) and, therefore, should be estimated simultaneously.

This same criticism can be applied to the approach of Hafley and Buford (1985). This new alternate methodology provides that link between the range and shape parameters. The methodology can be extended to the S_{BB} distribution, a bivariate distribution that is commonly used for describing simultaneously diameter and height structure (Tewari and Gadow 1999). In that case, the marginal distributions for diameter and height would be determined as outlined, but a strategy for determining the correlation parameter to tie the marginal distributions together would have to be developed.

The wealth of literature on diameter distribution modeling suggests no one system will universally fit all situations. In this study approximately 10 percent of the plots could not adequately be modeled using the S_B distribution. A better understanding of the biological relationships that underscore stand development over time is necessary to further develop existing approaches. Nevertheless, the S_B parameter recovery system developed in this paper is tractable for use in stand modeling and yielded good representations of unthinned loblolly pine stands.

Literature Cited

Ali, M.M. 1974. Stochastic ordering and kurtosis measure. Journal of the American Statistical Association. 69: 543-545.

Bailey, R.L.; Dell, T.R. 1973. Quantifying diameter distributions with the Weibull function. Forest Science. 19(2): 97-104.

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

Cao, Q.V.; Burkhart, H.E.; Lemin, Jr., R.C. 1982. Diameter distributions and yields of thinned loblolly pine plantations. FWS-1-82. Blacksburg, VA: Virginia Polytechnic Institute and State University. School of Forestry and Wildlife Resources. 62 p.

Clutter, J.L.; Bennett, F.A. 1965. Diameter distribution in old-field slash pine plantations. Rep. 13. Macon, GA: Georgia Forest Research Council. 9 p.

Clutter, J.L.; Fortson, J.C.; Pienaar, L.V. [and others]. 1983. Timber management: a quantitative approach. New York: John Wiley. 333 p.

Dell, T.R.; Feduccia, D.P.; Campbell, T.E. [and others]. 1979. Yield of unthinned slash pine plantations on cutover sites in the West Gulf region. Res. Pap. SO-147. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 84 p.

Draper, J. 1952. Properties of distributions resulting from certain simple transformations of the normal distribution. Biometrika. 39: 290-301.

Feduccia, D.P.; Dell, T.R.; Mann, Jr., W.F. [and others]. 1979. Yields of unthinned loblolly pine plantations on cutover sites in the west gulf region. Res. Pap. SO-148. U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 88 p.

Goodwin, E.T. 1949. The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(x)e^{-x^2}dx$. Proceedings of the Cambridge Philosophical Society. 45: 241-245.

Gove, J.H.; Patil, G.P. 1998. Modeling the basal area-size distribution of forest stands: A compatible approach. Forest Science. 44(2): 285-297.

Hafley, W.L.; Buford, M.A. 1985. A bivariate model for growth and yield prediction. Forest Science. 31(1): 237-247.

Hafley, W.L.; Schreuder, H.T. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research. 7: 481-487.

Hill, I.D.; Hill, R.; Holder, R.L. 1976. Algorithm AS 99. Fitting Johnson curves by moments. Applied Statistics. 25: 180-189.

Hyink, D.M. 1980. Diameter distribution approaches to growth and yield modeling. In: Brown, K.M.; Clarke, F.R., eds. Forecasting forest stand dynamics. Thunderbay, Ontario: Lakehead University, School of Forestry: 138-163.

Hyink, D.M.; Moser, Jr., J.W. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science. 29(1): 85-95.

Johnson, M.E.; Tietjen, G.L.; Beckman, R.J. 1980. A new family of probability distributions with applications to Monte Carlo studies. Journal of the American Statistical Association. 75: 276-279.

Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. Biometrika. 36: 149-176.

Kamziah, A.K.; Ahmad, M.I.; Lapongan, J. 1999. Nonlinear regression approach to estimating Johnson S_B parameters for diameter data. Canadian Journal of Forestry. 29: 310-314.

Kaplansky, I. 1945. A common error concerning kurtosis. Journal of the American Statistical Association. 40: 259.

Knoebel, B.R.; Burkhart, H.E. 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. Biometrics. 47(1): 241-253.

Matney, T.G.; Sullivan, A.D. 1982. Compatible stand and stock tables for thinned and unthinned loblolly pine stands. Forest Science. 28(1): 161-171.

Newbury, J.D.; Burk, T.E. 1985. S_B distribution-based models for individual tree merchantable volume-total volume ratio. Forest Science. 31: 389-398.

Popham, T.W.; Feduccia, D.P.; Dell, T.R. [and others]. 1979. Site index for loblolly plantations on cutover sites in the west gulf coastal plain. Res. Note SO-250. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 7 p.

Ralston, A.; Rabinowitz, P. 1978. A first course in numerical analysis. 2nd ed. New York: McGraw-Hill. 556 p.

Scolforo, J.R.S.; Thierschi, A. 1998. Estimativas e testes da distribuição de frequência diâmétrica para *Eucalyptus camaldulensis*, através da distribuição S_B, por diferentes métodos de ajuste. Scientia Forestalis. 54: 93-106.

Smalley, G.W.; Bailey, R.L. 1974. Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama, and Georgia highlands. Res. Pap. SO-96. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 81 p.

Tewari, V.P.; Gadow, K.V. 1999. Modelling the relationship between tree diameters and heights using S_{BB} distribution. Forest Ecology and Management. 119: 171-176.

Tham, Å. 1988. Estimate and test frequency distributions with the Johnson S_B function from stand parameters in young mixed stands after different thinning treatments. In: Ek, A.R.; Shifley, S.R.; Burk, T.E., eds. Forest growth modelling and prediction: Proceedings of the IUFRO Conference. Gen. Tech. Rep. NC-120. Minneapolis: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station: 255-262. Vol. 1.

Zarnoch, S.J.; Feduccia, D.P.; Baldwin, Jr., V.C.; Dell, T.R. 1991. Growth and yield predictions for thinned and unthinned slash pine plantations on cutover sites in the west gulf region. Res. Pap. SO-264. New Orleans: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 32 p.

Zhou, B.; McTague, J.P. 1996. Comparison and evaluation of five methods of estimation of the Johnson system of parameters. Canadian Journal of Forestry. 26: 928-935.

Zöhrer, F. 1972. The beta-distribution for best fit of stem-diameter-distribution. In: 3rd Conference advisory group of Forest Statisticians: Proceeding of the IUFRO Conference. INRA Publ. 72-3. Paris: Institut National Recherche Agronomique: 91-106.

Appendix

The partial derivatives of the moments of Y [$\mu_r'(y)$, see equation (8)] with respect to λ and δ are derived below. These derivatives are needed for the computation of the partial derivatives of equations (9) and(10). First, performing a change-of-variable from y_{median} to d_{median} [where we let *d*, for tree diameter, replace *x* in equation (4)], we have from equation (7) that $\gamma = \delta \ln\left[\lambda/(d_{\text{median}} - \xi) - 1\right]$. Substituting for γ where necessary, we have

$$
\begin{aligned}
\frac{\partial \mu_r'(y)}{\partial \lambda} &= -\frac{r}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(-z^2/2) \cdot \exp((\gamma - z)/\delta)}{[1 + \exp((\gamma - z)/\delta)]^{r+1}} \cdot D_\lambda \left[\frac{\delta \ln[\lambda/(d_{\text{median}} - \xi) - 1] - z}{\delta} \right] dz \\
&= -\frac{r}{\lambda - d_{\text{median}} + \xi} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(-z^2/2) \cdot \exp((\gamma - z)/\delta)}{[1 + \exp((\gamma - z)/\delta)]^{r+1}} \, dz \qquad \text{(A1)} \\
&= -\frac{r}{\lambda - d_{\text{median}} + \xi} \cdot \left[\mu_r'(y) - \mu_{r+1}'(y)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \mu_r'(y)}{\partial \delta} &= -\frac{r}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(-z^2/2) \cdot \exp((\gamma - z)/\delta)}{[1 + \exp((\gamma - z)/\delta)]^{r+1}} \cdot D_\delta \left[\frac{\delta \ln[\lambda/(d_{\text{median}} - \xi) - 1] - z}{\delta} \right] dz \\
&= -\frac{r}{\delta^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp(-z^2/2) \cdot \exp((\gamma - z)/\delta)}{[1 + \exp((\gamma - z)/\delta)]^{r+1}} \cdot z \, dz
\end{aligned}
$$

Because of the extra *z*, the integral does not reduce to the difference of two S_B moments as it did in the derivative with respect to λ. Using the pseudo-moment $\tilde{\mu}_r'(y)$ defined in equation (12), we can express the partial derivative as

$$
\frac{\partial \mu_r'(y)}{\partial \delta} = -\frac{r}{\delta^2} \cdot \left[\tilde{\mu}_r'(y) - \tilde{\mu}_{r+1}'(y)\right] \qquad \text{(A2)}
$$

Parresol, Bernard R. 2003. Recovering parameters of Johnson's S_B distribution. Res. Pap. SRS-31. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 9 p.

A new parameter recovery model for Johnson's S_B distribution is developed. This latest alternative approach permits recovery of the range and both shape parameters. Previous models recovered only the two shape parameters. Also, a simple procedure for estimating the distribution minimum from sample values is presented. The new methodology employs the median and the first and second moments of the distribution. The methodology is demonstrated by modeling diameter distributions of unthinned loblolly pine plantations. Compatible equations for projecting per-hectare values of number of trees and basal area from initial stand conditions are presented, as well as equations for predicting median diameter, mean diameter, and the location parameter. Given estimates of these five stand attributes, the range and the two shape parameters of the S_B distribution can be recovered. The χ^2 goodness-of-fit test rejected 56 cases out of 527 for conformance to an S_B distribution. Though the S_B distribution is very flexible in terms of distribution shape, about 10 percent of the loblolly plantation observations did not follow this distribution. Nonetheless, deviation analysis showed reasonable results, with 77 percent of the variation explained in current and projected distributions (numbers of trees by 2.5-cm diameter class). Overall, the recovered S_B distributions provided good approximations of the observed diameter distributions.

Keywords: Diameter distributions, kurtosis, Newton-Raphson method, moments, parameter recovery model, percentiles, *Pinus taeda*, skewness.