BIOMETRICS *33*, 471–478
September, 1977

A Useful Bivariate Distribution for Describing Stand Structure of Tree Heights and Diameters

H. T. SCHREUDER

USDA Forest Service, Southeastern Forest Experiment Station, Forestry Sciences Laboratory, Research Triangle Park, North Carolina 27709, U.S.A.

W. L. HAFLEY

School of Forest Resources, North Carolina State University, Raleigh, North Carolina 27607, U.S.A

Summary

Hafley and Schreuder (1977) have shown that the marginal S_B distribution fits diameter and height data consistently better than the Weibull, beta, gamma, lognormal, and normal distributions. The bivariate extension of the S_B distribution, the S_{BB}, is both more realistic and provides more usable information than the currently accepted approach in describing even-aged forest stand height-diameter data. The S_{BB} allows for the generation of bivariate frequencies for diameter and height, whereas the current approach only provides marginal frequencies for diameter. In addition, the S_{BB} implies a new height-diameter relationship which is comparable in fit to the most commonly used height-diameter regression model. Application of the S_{BB} to two data sets is presented.

1. Introduction

An important problem in forestry is the prediction of stand yields in volume on the basis of stand age, productive capacity of the site, and stand density. Forest managers are interested in studying the effect of such stand management practices as thinning and fertilization on volume in pulpwood, sawtimber, and veneer. The volume in each of these products is heavily dependent on tree diameter and height distributions. For example, a given stand volume consisting of only small trees may be entirely pulpwood; whereas, the same volume in a few large trees may be primarily sawtimber. Hence, there is considerable interest in the successful fitting of a bivariate statistical distribution to describe diameter-height frequency data. We are not aware of such a successful application in the literature. Current practice is to fit a marginal distribution to the diameter frequency data and then use an empirical height-diameter relationship to estimate average height per diameter class and thence volume (see Clutter and Allison 1974). Although this latter approach is satisfactory in many ways, it ignores the fact that height can vary considerably for a given diameter. This is especially important in older, commercially more important stands. Hence, the possibility of generating a bivariate distribution of diameters and heights for a stand would often be of interest to managers.

Key Words: Bivariate tree height-diameter distribution; Even-aged stands; Median regression; Fitting parameters; Maximum likelihood; Johnson distributions; S_{BB}.

In an earlier paper (Hafley and Schreuder 1976) several bivariate distributions potentially useful for describing the joint frequency distribution of tree diameters and heights in even-aged stands of timber were reviewed. None of the available or derived bivariate generalizations of the univariate lognormal, gamma, and Weibull distributions provided reasonable height-diameter relations.

In a recent paper (Hafley and Schreuder 1977) the S_B distribution (Johnson 1949a) was introduced to the forestry literature. This distribution fitted the marginal frequencies of both diameter and height consistently better than the Weibull, beta, gamma, lognormal, and normal distributions. A bivariate generalization of the S_B, the S_{BB} (Johnson 1949b), implies a relationship between height and diameter in even-aged stands which seems to fit data as well as regression models now widely accepted in the forestry literature.

In this paper we present a discussion of the S_{BB} distribution and some of its relevant properties, show briefly how to estimate the parameters, and give examples of its use for forestry height-diameter data sets.

1.1 The Bivariate S_B Distribution (S_{BB})

The S_{BB} distribution (Johnson 1949b) is a bivariate distribution for which both marginals are S_B distributions. It is the bivariate distribution of y_1 and y_2 when the standard normal variates z_1 and z_2 are defined as

$$z_1 = \gamma_1 + \delta_1 \ln \{y_1/(1 - y_1)\} \quad \text{and} \quad z_2 = \gamma_2 + \delta_2 \ln \{y_2/(1 - y_2)\}, \tag{1}$$

where z_1 and z_2 have the joint normal bivariate distribution with correlation ρ; namely,

$$p(z_1, z_2; \rho) = [2\pi\sqrt{1 - \rho^2}]^{-1} \exp \{-(1/2)(1 - \rho^2)^{-1}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\}.$$

For our purposes $y_1 = (D - \xi_1)/\lambda_1$ and $y_2 = (H - \xi_2)/\lambda_2$ where ξ_1 and ξ_2 are the smallest values and λ_1 and λ_2 are the range of diameter (D) and height (H), respectively, in the population. Many of the properties of the S_{BB} can be obtained directly from the literature of the bivariate normal distribution because of the relation between them.

One of the properties of interest is the regression relation between y_2 and y_1. In general, the usual mean regression is complicated. However, the median regression takes a much simpler form; namely,

$$y_2 = \theta y_1^{\phi}\{(1 - y_1)^{\phi} + \theta y_1^{\phi}\}^{-1} \tag{2}$$

where $\theta = \exp \{\rho\gamma_1 - \gamma_2)/\delta_2\}$ and $\phi = \rho\delta_1/\delta_2$ ($\phi > 0$ since we assume $\rho > 0$). For $0 < \phi < 1$, the first derivative of y_2 with respect to y_1 does not exist at the extremes, $y_1 = 0$ and $y_1 = 1$, and thus may not be of practical use. For data sets of diameter and height the constraint $\phi > 1$ should be imposed on parameter estimation. The regression is linear if $\rho\delta_1 = \delta_2$ and $\rho\gamma_1 = \gamma_2$.

Rewritten in terms of diameter and height, equation (2) can be expressed as

$$(H - \xi_2)/\lambda_2 = \theta\left[\left(\frac{\xi_1 + \lambda_1 - D}{D - \xi_1}\right)^{\phi} + \theta\right]^{-1} \tag{3}$$

which does not reduce to a more easily interpretable model.

A second property of interest is the fact that the conditional distribution of y_2 given y_1 also follows the S_B distribution with parameters γ' and δ' where

$$\gamma' = [\gamma_2 - \rho\{\gamma_1 + \delta_1 f_B(y_1)\}](1 - \rho^2)^{-1/2}$$

and

$$\delta' = \delta_2(1 - \rho^2)^{-1/2},$$

with

$$f_B(y_1) = \ln \{y_1/(1 - y_1)\}.$$

This conditional distribution can be bimodal; however, unimodality is assured if $\rho^2 > 1 - 2\delta^2_2$. From the conditional distribution we can obtain percentile limits for the median regression. The α-percentile for y_2 given y_1 is obtained by solving for y_2 in the relation

$$z_\alpha = [\gamma_2 + \delta_2 f_B(y_2) - \rho\{\gamma_1 + \delta_1 f_B(y_1)\}](1 - \rho^2)^{-1/2} \tag{4}$$

where z_α is the α-percentile of the standard normal deviate. In terms of equation (3), the α-percentile of H, H_α, can be obtained from

$$(H_\alpha - \xi_2)/\lambda_2 = \psi(1 + \psi)^{-1} \tag{5}$$

where

$$\psi = \theta\left(\frac{D - \xi_1}{\xi_1 + \lambda_1 - D}\right)^\phi \exp\,[z_\alpha\sqrt{1 - \rho^2/\delta_2}].$$

1.2 Fitting the Distribution

To fit the S_{BB} distribution, one first obtains estimates of the parameters of the marginal S_B distributions, transforms the observation pairs to standard normal deviates, and finally, estimates the correlation between these standard normal deviates (Johnson 1949b). The S_B distribution does not have closed solution maximum likelihood estimators (MLE) for the four parameters, ξ, λ, γ, and δ, and requires an iterative estimation procedure. However, if ξ and λ are given, closed solution MLE of γ and δ do exist. These we shall refer to as conditional MLE. While it is not always easy to specify *a priori* values for both ξ and λ, a reasonable value for ξ is often obvious, reducing the iteration effort.

For the diameter-height relationship, it is simplest and often reasonable to specify values of ξ_1 and ξ_2 of 0 and 4.5, respectively. Since diameters are measured at 4.5 feet (1.37 m) above ground level by convention, it is desirable that any height-diameter regression pass through this point. Further, it seems intuitive that $(\xi_i + \lambda_i)$, $i = 1, 2$, must be larger than the largest respective observation in the data set. Thus, we iteratively search for the MLE of λ_i using the conditional closed solution MLE of γ_i, and δ_i, such that $\xi_i + \hat{\lambda}_i$ is larger than the largest observed appropriate value, diameter or height. The conditional MLE of γ_i and δ_i are computed from

$$\hat{\gamma}_i = -\bar{f}_i/s_i \tag{6}$$

and

$$\hat{\delta}_i = 1/s_i, \tag{7}$$

where

$$\bar{f}_i = \left(\sum_{j=1}^{n} f_{ij}\right)\Big/n; \qquad s_i^2 = 1/n \sum_{j=1}^{n} (f_{ij} - \bar{f}_i)^2; \qquad f_{ij} = \ln \{y_{ij}/(1 - y_{ij})\};$$

$$y_{ij} = (x_{ij} - \xi_i)/\hat{\lambda}_i, \quad \text{with} \quad x_{1j} = D_j \quad \text{and} \quad x_{2j} = H_j, \qquad i = 1, 2; \quad j = 1, \cdots, n;$$

and n is the number of height-diameter pairs in the data.

Hahn and Shapiro (1967) indicate that the quality of fit is relatively unaffected by the choice of lower bound and range so long as they are consistent with the data to be fitted. But, one pays a price for getting a good fit with quite different values of the parameters. And, this price must be high correlation between the parameter estimates. Hence, one must make sure that the estimates of ξ and λ are reasonable values based on knowledge of the data.

The only additional variable to estimate in fitting the S_{BB} distribution is the correlation coefficient ρ. This is obtained from

$$\hat{\rho} = \sum_{j=1}^{n} z_{1j} z_{2j}/n, \tag{8}$$

where $z_{ij} = \hat{\gamma}_i + \hat{\delta}_i \ln \{y_{ij}/(1 - y_{ij})\}$, $i = 1, 2; j = 1, \cdots, n$, and y_{ij} is as above.

As indicated earlier, values of $\phi < 1$ $[\phi = \rho\delta_1/\delta_2]$ are unacceptable for the height-diameter relationship. When such values occurred in our data sets we incremented $\hat{\lambda}_1$, the estimate of the diameter range parameter, by the width of a diameter class until $\phi > 1$ was attained. Since the H-D regression terminates at the point $(\xi_2 + \hat{\lambda}_2, \xi_1 + \hat{\lambda}_1)$, incrementing $\hat{\lambda}_1$ sufficiently while holding $\hat{\lambda}_2$ constant will eventually cause ϕ to be greater than 1. The obvious implication of an estimated $\phi < 1$ is that the data do not support the assumption of an S_{BB} distribution using standard fitting procedures. Since we feel the relationship implied by $\phi < 1$ is not realistic, we have imposed a constraint on the fitting process to avoid such cases. In the

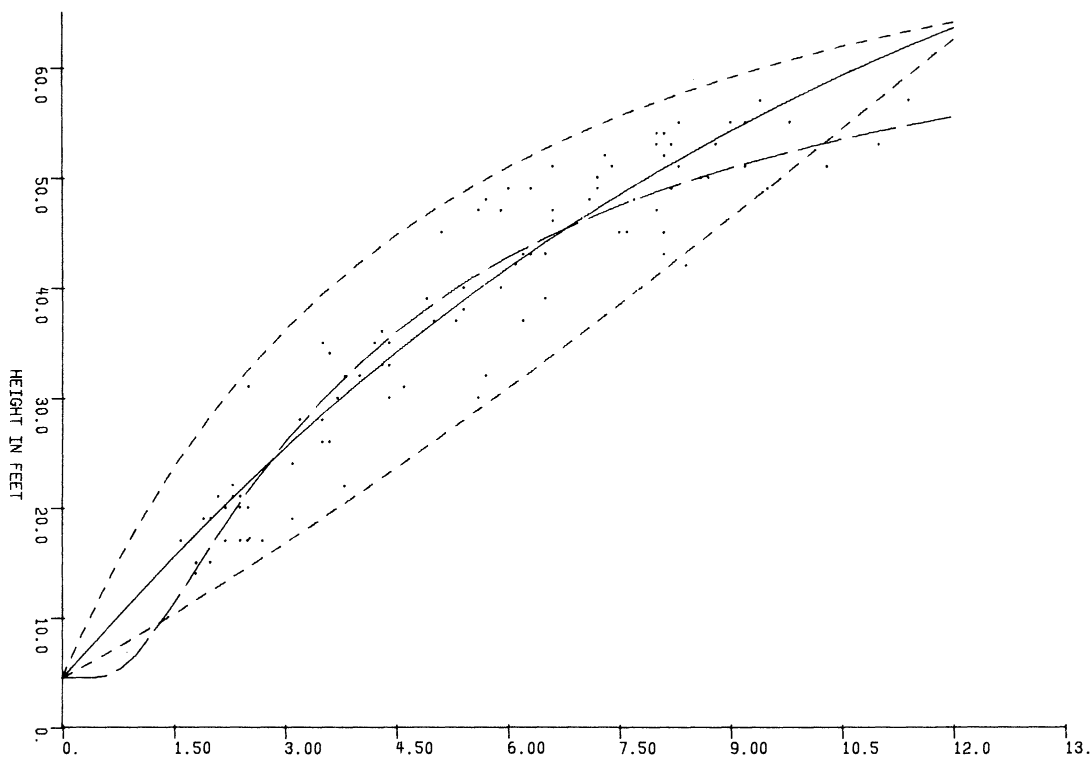

Figure 1.

Scatter plots of the observed data points with the standard regression (long dash line) and the S_{BB} regression (solid line) with its five- and 95-percentile lines (short dash lines) superimposed for stand 1.

21 data sets we examined, imposition of this constraint was necessary in only six sets, and in no case was it necessary to increase $\hat{\lambda}_1$ more than three diameter classes.

2. *Results*

The S_{BB} distribution was fitted to five plantation and 16 natural stand data sets, each consisting of a large number of diameter and height measurements on one-acre (0.40 ha) plots. Sample sizes ranged from 105 to 728 trees. However, only the results for two representative data sets are presented here. They consist of 105 and 566 trees, respectively, measured in 20- and 18-year-old natural shortleaf pine (*Pinus echinata* Mill.) stands located in northern Florida. Diameters were measured to the nearest 0.1 inch (0.25 cm) and heights to the nearest foot (30 cm).

In Figures 1 and 2 we present scatter diagrams of the two data sets; the median regression line from the S_{BB} fit with the five- and 95-percentile lines included and a regression line based on the regression equation

$$\ln (H - 4.5) = a + b/D \tag{9}$$

fit to the same data are superimposed. The model (9) is used frequently by practicing

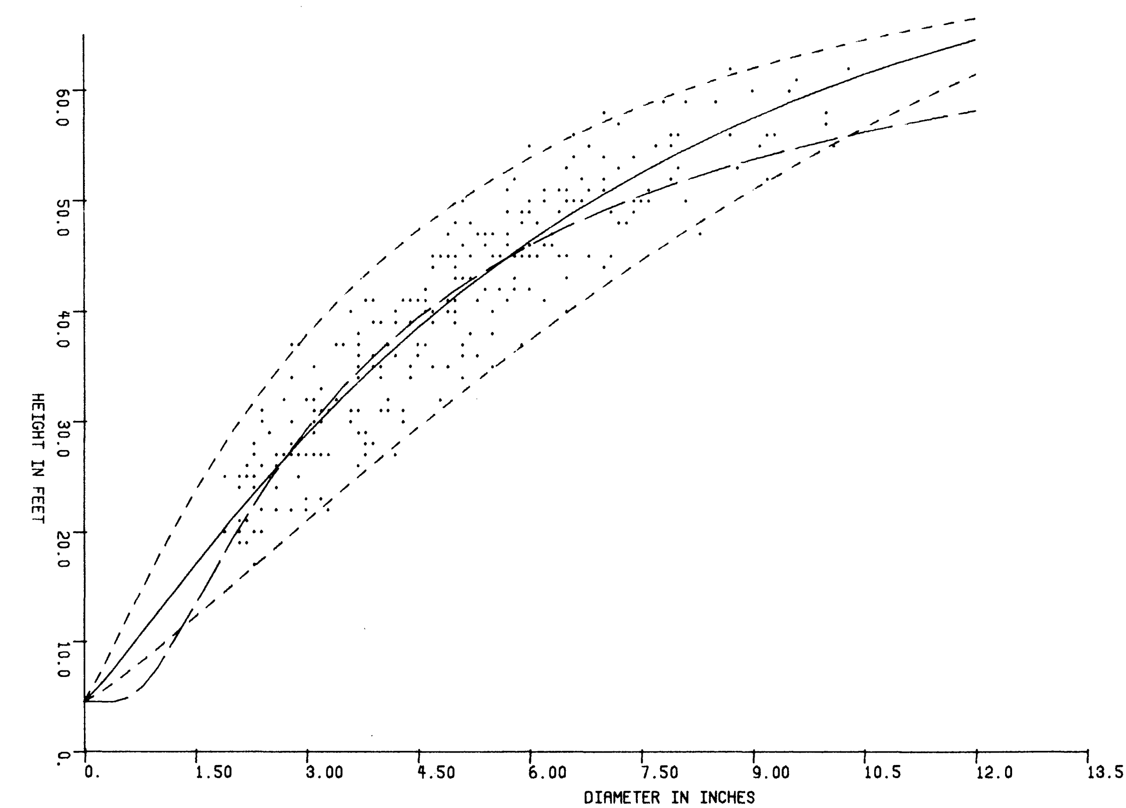

Figure 2

Scatter plots of the observed data points with the standard regression (long dash line) and the S_{BB} regression (solid line) and its five- and 95-percentile lines (short dash lines) superimposed for stand 2.

foresters and has been shown to perform well by statistical standards (Curtis 1967). The equations of the regression lines are presented in Table 1 along with the actual deviations, as plotted in Figures 1 and 2, compared on the basis of squared, absolute, and simple deviations. In Table 2, the observed and expected joint probabilities are shown.

While it would be possible to find an empirical model that would fit the data better than (9), we have used (9) because of its wide acceptance in forestry.

3. Discussion

In Hafley and Schreuder (1976) we evaluated the usefulness of bivariate distributions for describing height-diameter data based on the criteria that the distribution should give satisfactory fits to both the marginal frequencies of heights and diameters and also provide a regression relationship in accordance with forestry literature (for example Curtis 1967). Although some of the bivariate distributions described there, primarily the bivariate Weibull distributions, can give satisfactory fits to marginal distributions of diameters and heights, none indicated an acceptable height-diameter relationship.

In Hafley and Schreuder (1977) we demonstrated the appropriateness and superiority of the S_B over the Weibull and beta in describing the marginal frequency distributions of diameter and height. To evaluate the S_{BB} further, note that the shape of the regression line given $\phi > 1$ satisfies existing knowledge of the general shape of the relationship between height and diameter (Figures 1 and 2). The regression line based on the S_{BB} starts at the estimated lower bounds and ends at the estimated extremes of the marginal distributions. The plot of the five- and 95-percentile lines obtained by substituting the MLE into equation (4) reasonably approximates the heteroscedasticity of the data. This is in conformance with one hypothesis for even-aged stands of pine that the tallest (dominant) trees and the shortest (suppressed) trees are associated, respectively, with the largest and smallest diameters, and vary little in height within their respective range of diameters. Tree diameters more closely associated with the mean diameter (codominant and intermediate trees) of the stand exhibit much more variation in height. All 21 data sets in Hafley and Schreuder (1977) tend to support this hypothesis.

The measures of fit shown in Table 1 provide confirmation of the visual impression one

Table 1

Comparison of the standard and S_{BB} regression curves in terms of the sums of squares (SR), absolute (AD), and simple (D) deviations from regression for stands 1 and 2.

Data set	Equation	Regression	SR	AD	D
Stand 1	Standard	$\ln(H - 4.5) = 4.22 - 3.46/D$	2,203.48	377.62	65.14
	S_{BB}	$H - 4.5 = 104.81\left[\left(\frac{12.4 - D}{D}\right)^{1.03} + 1.74\right]^{-1}$	2,150.53	366.81	-15.61
Stand 2	Standard	$\ln(H - 4.5) = 4.24 - 3.09/D$	12,934.12	2,136.18	160.90
	S_{BB}	$H - 4.5 = 196.67\left[\left(\frac{15.4 - D}{D}\right)^{1.14} + 3.03\right]^{-1}$	11,922.26	2,061.09	-63.91

gets from Figures 1 and 2 that the S_{BB} regression gives a slightly better fit than the standard. The standard seems to fit the lower range of the data well but tends to underestimate the upper end.

The primary advantage of the S_{BB} approach is that it allows for the generation of the bivariate frequencies of height and diameter. While initially the use of the S_{BB} will be of most value in simulation activities, the forest manager is interested in the distribution in size of both height and diameter. The S_{BB} could be a useful tool for gaining that information.

The results in Table 2 show that the calculated and observed frequencies are similar in the main part of the data range but are different in the tail regions. These differences should not be surprising because the sample sizes, although relatively large, are still small for adequately fitting tail probabilities of bivariate distributions. By statistical standards the results in Table 2 for the marginal distributions are satisfactory as measured by the Kolmogorov-Smirnov goodness-of-fit test. Calculated d-values for diameter and height are $d = 0.087$ and $d = 0.076$ for stand one (tabulated $d = 0.133$ at the five percent level) and $d = 0.028$ and $d = 0.044$ for stand two (tabulated $d = 0.057$ at the five percent level) indicating satisfactory fits (Massey 1951).

Our experience with bivariate distributions thus far indicates that the S_{BB} shows the most promise in describing diameter-height data. The close connection of the S_{BB} to the bivariate normal makes many of the properties and computational aids for the latter applicable to data

Table 2

Observed and predicted (O/P) frequencies for height-diameter classes for stands 1 and 2[1].

Diameter (inches)	Height (feet) 15	20	25	30	35	40	45	50	55	60	65	70
						Stand 1						
2	8/2.5	11/2.8	0/1.6	1/0.6								
3	1/0.8	1/2.8	2/3.6	2/2.7	1/1.3	0/0.4						
4		1/1.0	2/2.7	4/3.9	6/3.5	0/1.9	0/0.6					
5			0/1.0	1/2.6	2/4.0	3/3.8	1/2.2	0/0.6				
6				2/1.0	1/2.5	3/4.1	7/4.0	3/2.1				
7					0/0.9	0/2.4	3/4.1	7/3.9	1/1.4			
8						1/0.8	4/2.3	4/4.2	6/5.2	2/0.4		
9							0/0.6	4/2.2	4/4.0	0/1.3		
10								1/0.4	1/2.1	0/2.5	1/0.1	
11									2/0.2	0/1.6	0/0.2	
						Stand 2						
2	12/6.9	33/6.4	26/2.2	0/0.3								
3	1/0.1	11/3.5	32/16.8	30/23.4	15/15.0	2/5.1	0/0.9	2/0.1				
4		1/0.2	10/3.7	36/18.4	38/35.2	30/33.4	0/17.0	2/4.4	0/0.5			
5			0/0.2	7/2.5	20/13.4	35/32.9	45/41.7	6/27.4	0/8.1	0/0.8		
6				2/0.1	3/1.3	15/7.9	41/23.9	41/37.7	6/28.4	3/7.7	0/0.4	
7						2/0.6	9/4.0	31/15.2	15/28.3	4/20.8	0/3.3	
8						1/0.0	6/0.2	14/1.9	20/9.1	10/18.5	2/9.3	0/0.2
9							2/0.0	6/0.1	14/0.9	6/5.6	0/9.4	0/1.2
10								1/0.0	5/0.0	6/0.5	1/3.3	0/1.9
11									2/0.0	0/0.0	0/0.3	0/1.0

[1] The differential between the observed totals and predicted totals (105 versus 99.4 and 566 versus 563.5) is contained primarily in the one-inch diameter class.

satisfying the S_{BB} distribution. Further, application of the S_{BB} results in a consistent bivariate height-diameter system which could conceivably be generalized to the even more desirable volume-height-diameter trivariate system. In the burgeoning activity of forest simulation, these increases in realism and consistency are desirable characteristics.

Acknowledgment

We appreciate Dr. N. L. Johnson's many suggestions without which this paper would not have been possible.

Une distribution bivariate utile pour décrire la structure d'un peuplement forestier en hauteurs et diamètres.''

Résumé

Une distribution bivariate S_{BB} est décrite; elle est à la fois plus réaliste et plus utile que celles que l'on utilise couramment pour décrire des données hauteur-diamètre de peuplements forestiers d'âge constant. A l'aide de S_{BB} on peut générer des fréquences bivariates pour diamètre et hauteur; alors que l'approche habituelle ne fournit que les fréquences marginales des diamètres. A cela s'ajoute que la S_{BB} implique une nouvelle relation hauteur-diamètre dont l'ajustement est comparable à celui utilisé le plus couramment pour le modèle de regression hauteur-diamètre. Hafley et Schreuder (1977) ont montré que la distribution marginale S_B permet un ajustement sensiblement meilleur des données de diamètre et de hauteur que les distributions de Weibull, bêta, gamma, lognormale et normale. One présente l'application de S_{BB} à deux ensembles de données.

References

Clutter, J. L. and Allison, B. J. (1974). A growth and yield model for *Pinus radiata* in New Zealand. In *Growth Models for Tree and Stand Simulation.* J. Fries (ed), Royal College of Forestry, Stockholm, Sweden. Department of Forest Yield Research Note 30, 136–160.

Curtis, R. O. (1967). Height-diameter and height-diameter-age equations for second-growth Douglas Fir. *Forest Science 13(4),* 365–375.

Hafley, W. L. and Schreuder, H. T. (1976). Some non-normal bivariate distributions and their potential for forest application. *International Union Forest Research Organizations. XVI World Congress Proceedings, Div. VI,* 104–114. (Oslo, Norway; June 20-July 2, 1976).

Hafley, W. L. and Schreuder, H. T. (1977). Statistical distributions for fitting diameter and height data in even-aged stands. (Accepted by *Canadian Journal of Forest Research.*)

Hahn, G. J. and Shapiro, S. S. (1967). *Statistical Models in Engineering.* John Wiley and Sons, Inc., New York.

Johnson, N. L. (1949a). Systems of frequency curves generated by methods of translation. *Biometrika 36,* 149–176.

Johnson, N. L. (1949b). Bivariate distributions based on simple translation systems. *Biometrika 36,* 297–304.

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association 46,* 68–78.

Received November 1976; *Revised March* 1977