A SIMPLIFIED METHOD FOR QUANTIFYING DIAMETER DISTRIBUTIONS OF NORTHERN MICHIGAN UPLAND HARDWOODS WITH THE WEIBULL AND S_B FUNCTIONS

By

Robert De Geus

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Forestry

1986

ABSTRACT

A SIMPLIFIED METHOD FOR QUANTIFYING DIAMETER DISTRIBUTIONS OF NORTHERN MICHIGAN UPLAND HARDWOODS WITH THE WEIBULL AND S_B FUNCTIONS

by

Robert De Geus

This study presents a unified and simplified method for applying the Weibull or S_B distribution to modelling of tree diameter distributions. This approach uses spreadsheet software and a microcomputer rather than a mainframe computer.

The methodology is based on the use of point sample data and grouped frequencies of observed data. Grouped frequencies, and maximum likelihood and percentile estimation are used to estimate distribution parameters. The Weibull and S_B functions are compared for both estimation methods.

Ten ecological groupings of 72 upland hardwood stands are used for diameter distribution modelling. The stands are composed of mixed species and age groups.

The methods described are found to provide reasonably good models. The Weibull function with maximum likelihood parameter estimation is found to perform best with these methods and data. Percentile parameter estimation for the Weibull is suggested as the starting point with these methods for obtaining the maximum likelihood estimates.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Carl Ramm, graduate committee chairman, for his belief in a light hand on the reins. It is he, Dr. Kurt Pregitzer, and Dr. David Lusch that have made completion of this thesis possible.

Recognition is due Niro Higuchi of INPA, Manaus, Brazil for providing perspective in an unfamiliar environment. The education represented by this thesis would not have been complete without him.

Dr. Phu Nguyen deserves many thanks for his help and friendship. Steve Westin and George Host were invaluable in many ways.

The largest share of gratitude is reserved for my wife, Reenie, as well as my family. They must be wizards because I feel like the scarecrow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Tree diameters are of interest to forest managers and researchers for a variety of reasons. First and foremost is the fact that tree diameter is well correlated with other important variables such as volume, value, conversion cost, and product specification. Secondly, from a silvicultural standpoint, diameter distributions are useful in understanding current and future stand structure.

Obtaining a statistical or mathematical model of an observed diameter distribution can be an important step in developing growth and yield projection models. Recent projection models have incorporated diameter distribution models to obtain better projection accuracy. The improvement in projection is related to the capacity of a projection model to project stand attributes by diameter class. By predicting diameter distributions at a given time from stand variables, stand simulators using such models maintain some of the detail of individual tree simulators without the necessity for keeping track of individual trees throughout the simulated rotation (Little, 1983). This distributional information is useful in delineating yield by diameter class, identifying silvicultural opportunities, and projecting the impact of management decisions over a rotation.

Distributional models have not been developed for forestry. Instead, general distribution functions that can generate a wide variety of curve shapes have been adapted for use in modelling tree diameter distributions. Any distribution model is dependent upon certain constants or parameters for use. These parameters are estimated from sample data. For the normal distribution, these parameters (mean and standard deviation) are familiar and well understood. The parameters of other distributions, such as those used for modelling diameter distributions, are not well understood. The result is that parameter estimation is often a difficult or vague process.

One of the more important distribution models in tree diameter distribution modelling is the Weibull distribution. The Weibull distribution has been successfully applied to many diameter distribution modelling problems, principally in even-aged, single species plantations or forests. Another model that is relatively new to diameter distribution modelling is the S_B distribution of N.L. Johnson (1949). The S_B distribution is more complex than the Weibull and is theoretically capable of producing a wider range of curve shapes (Hafley and Schreuder, 1977).

The Weibull distribution is widely accepted in tree diameter distribution modelling and in growth and yield modelling based on distribution models. There is, however, no biological basis for the use of this or any distribution model in forestry, and there remains some uncertainty about model selection and application (Bailey, 1980). Therefore, models such as the S_B are of interest in the search for the best possible model.

The scope of this study is two-fold. First, to address the problem of non-normal distribution parameter estimation, a simplified (and as explicit as possible) methodology for parameter estimation for the Weibull and S_B distributions is described and applied. Second, the relative performance of the Weibull and S_B distributions are assessed using observed diameter distributions from ten ecological forest groupings. The purpose of the first objective is to establish parameter estimation and distribution derivation methods that facilitate diameter distribution modelling while maintaining as much accuracy as possible. As such, three new facets of diameter distribution modelling are presented. First, data obtained from sampling with probability proportional to size (point sampling) are used to derive observed diameter distributions and estimate distribution parameters. Second, grouped frequency counts are used for parameter estimation. Third, mainframe computing is avoided in favor of more accessible (and cheaper) microcomputers with spreadsheet software. In addition, the data used comprise mixed species and age upland hardwoods. This latter point is a function of the data available as well as an interest in applying simplified methods to complicated data.

The purpose of the second objective is to investigate the relative performance of two distribution models using two parameter estimation methods on observed data. This investigation assesses the different models as well as their underlying methodologies.

The ultimate goal is the application of the appropriate model to growth and yield modelling with the observed data. This goal is not addressed in this study. However, the results of the study include distribution parameter estimates that may be appropriate for future studies. Finally, the methods presented may serve as a basis for future diameter distribution modelling either as the current data base develops or for an entirely new data set.

Literature Review

I. Background

DeLiocourt first introduced mathematical and statistical models for tree diameter distributions to forestry in 1898. Since that time, a variety of models have been investigated for application to tree diameter distributions such as the negative exponential, negative power function, gamma, beta, lognormal, Weibull, and Johnson's S_B (Hafley and Schreuder, 1977). The goal in these studies was to find a model that could adequately describe diameter distributions based on observed sample data. The desirable properties of such a function were given by Hafley and Schreuder (1977):

1. the model should have a mathematical form that allows ease of computation;
2. the parameters of the function should have properties that make their estimation relatively simple and exact;
3. the function should have the capacity to generate the widest possible range of curve shapes without sacrificing accuracy of fit to observed data.

The last feature given was related to the fact that observed diameter distributions take on a variety of shapes. These shapes ranged from a reversed J-shaped curve in all-aged, undisturbed forests to a mound shaped, negatively skewed curve in mature conifer plantations (Bailey and Dell, 1973; Lorimer and Krug, 1980). Bailey (1980) further suggested that a single function or distributional family is desirable for describing all possible distributional curves for tree diameters in a forest. The single function approach was shown to accomodate growth and yield projection.

With regard to the models mentioned above, Hafley and Schreuder (1977) noted that the beta distribution and Johnson's S_B distribution yield the wider range of curve shapes. They based their conclusion on the plot of skewness (B_1) and kurtosis (B_2) values possible for each distribution function (Figure 1). Within this graph was an "impossible region" wherein no combinations of skewness and kurtosis could mathematically exist. According to their plot, the normal distribution occupied a single point, implying a single possible shape. The Weibull, gamma, lognormal, and exponential distributions each were represented by a line. The beta distribution and Johnson's S_B distribution each occupied a region within the graph, implying more flexibility of shape than a line alone.

Of those distributions that have been investigated, Hafley and Schreuder (1977) noted that only the Weibull and Johnson's S_B functions meet all the criteria given for an appropriate distribution for modelling tree diameter distributions. The other functions (i.e., the gamma,

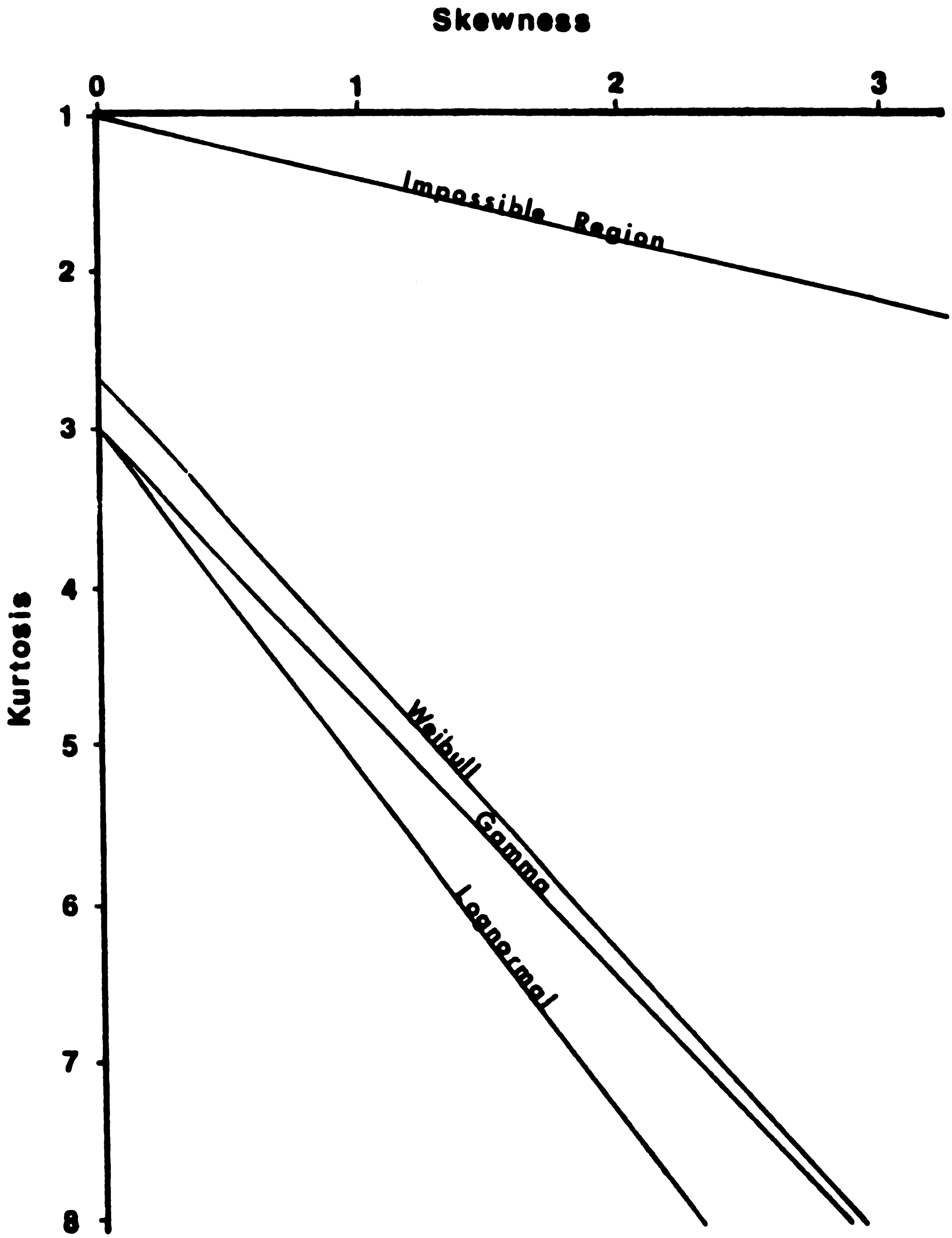

Figure 1: Skewness- Kurtosis space

lognormal, exponential, and beta distributions) showed problems with relative inflexibility, difficulty in parameter estimation, or difficulty in mathematical use of the function.

In general, the mathematical form of a distribution was given by all authors as the probability density function (pdf) denoted as f(x). A further criterion for a pdf was given by Johnson (1949) as ease of integration of f(x) to obtain F(x), the cumulative probability density function (cdf). The cdf was given as a means of determining the proportion of observations that would occur below a certain threshold observation.

In all studies citing criteria for distribution selection, the emphasis was on simplicity and/or ease of calculation in the use of a model. Most literature has not been concerned with this feature and Johnson (1949) claimed that this has held back the practical application of non-normal distributions. The result has been that the dominant distribution and theory in statistical applications has been the normal distribution (Johnson, 1949). One major reason given for this dominance was the ease with which moments of the distribution and, therefore, distribution parameter estimates can be obtained.

Johnson (1949) noted that three basic approaches to distribution parameter estimation exist. These were given as maximum likelihood estimation, the method of moments, and percentile estimation. He also summarized two conceptual approaches to the use of distribution models in fitting observed frequencies. The first was the idea that there should exist a theoretical, statistical basis in applying a distribution function to observed data. This idea was further expanded to state that the parameters of a function must be estimated on the basis of statistical theory in order to ensure that these estimates accurately represent the underlying population distribution parameters. The second concept was that the most important criterion in using a distribution model to fit observed frequencies is nothing more than the goodness of the fit. Any underlying theory or statistical parameter estimation was said to be of little interest for this approach. Johnson (1949) concluded that subscribing to either approach is dependent on the individual's objectives in using a distribution model. He suggested that a desire to develop inferences about an underlying population based on the use of a distribution model would require some degree of theoretical basis.

Maximum likelihood parameter estimation was said by Cohen (1965) to provide unique estimates of parameters with minimum bias in a statistically sound manner. He showed that maximum likelihood methods are based on the distribution function itself as applied to observed frequency data. Johnson (1949) stated that the method of moments is most practically applied to the normal distribution since calculation of sample moments for non-normal distributions is often very difficult. As such, the method of moments was not given as a general procedure. Percentile estimation was presented as something of the pattern maker's approach to curve fitting. The generalized method was given by Dubey (1967) as one in which observations at particular percentile points of the observed distribution are used to estimate parameters of a curve identical to the observed curve. Percentile parameter estimation was noted to yield estimates that are not unique and may not have much theoretical relation to the underlying population distribution.

Johnson (1949) related the methods of parameter estimation to the differing conceptual approaches to distribution function use. He wrote that from a theoretical standpoint, parameter estimates ought to be unique and should have some theoretical relationship to the underlying population distribution. From the standpoint of curve fitting alone, parameter estimates were not required to be unique or theoretically related to some underlying population distribution. In this latter case, Johnson (1949) concluded that any inference about a population using a distribution model would be entirely dependent upon the sampling method used and the depth of inference desired. The former conceptual approach was given as being appropriate for maximum likelihood parameter estimation. The latter case was given as being appropriate for either method of estimation.

A generalized approach to the application of distribution models to observed data was given by Johnson and Kotz (1970). They showed that observations on a continuous variable could be grouped into strata or classes, usually of equal width. They suggested that while the resulting variable was then discrete, this transformation was one merely of convenience. They wrote that inference or graphic presentation could be done on the basis of the original continuity of the data.

II. The Distribution Models

A. The Weibull Distribution

Weibull (1951) introduced a distribution model that gained wide acceptance in materials strength testing. Since its introduction, many studies have explored other applications and the intricacies of the distribution's properties. Despite these investigations, Dubey (1967) claimed that the Weibull distribution is not well understood.

Bailey and Dell (1973) applied the Weibull distribution to the problem of modelling tree diameter distributions. Since their introduction, the Weibull distribution has gained a great deal of attention and, to some degree, acceptance in forestry. There was no biological basis for their application; Bailey and Dell (1973) wrote that the Weibull distribution was chosen because of the relatively wide range of curve shapes it produces, the relative ease with which parameters may be estimated, and the ease by which the pdf can be integrated to obtain F(x). They found that the Weibull distribution effectively modelled even-aged conifer plantation diameter distributions.

Bailey and Dell (1973) gave the three-parameter Weibull probability density function as

$$f(x) = (c/b)(((x-a)/b)\hat{}(c-1)) \exp[-((x-a)/b)\hat{}c]$$

where x = diameter at breast height (dbh)

a = location parameter

b = scale parameter

c = shape parameter.

The location parameter (a) was described by Zarnoch, Ramm, Rudolph, and Day (1980) as the smallest diameter in the population. In general, it was given by Dubey (1967) as the lower end-point of the distribution. Thus, in a population of trees, the location parameter (a) was noted to be theoretically zero.

The scale parameter (b) was described by Little (1983) as the 63rd percentile of a population. In other words, b is the diameter below which 63% of all ordered observations would occur.

The shape parameter (c) was said by Bailey and Dell (1973) and Lorimer and Krug (1983) to be associated with the general degree of skewness of the distribution. They showed that for $c<1$, the curve becomes a reversed J-shape, or highly skewed in the positive direction. For $1<c<3.6$ the curve exhibits a mound shape with positive skewness; for $c=3.6$ the curve approximates a normal curve; when $c=1$ the exponential distribution results; for $c>3.6$ the curve becomes negatively skewed. As c approaches infinity, the curve becomes a spike over a single point.

Integration of the pdf resulted in the cumulative probability function (Bailey and Dell, 1973):

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

with parameters as given above.

This function was shown to provide the percentage of observations below a given diameter (x). Johnson and Kotz (1970) mentioned that in a strict sense it is necessary to use F(x) to obtain percentages for classes formed from a continuous variable. They further stated, though, that the probability density function f(x) may be used to obtain percentages for classes without any significant loss of accuracy in fit. In such a case they suggested that inferences based on such an approach could suffer some loss of precision because the continuous variable is treated altogether as a discrete variable.

The moments of the Weibull distribution were presented by Johnson and Kotz (1970). It was apparent that under any circumstance the moments present serious difficulty in calculation.

Estimation of the parameters of the Weibull distribution was shown by Cohen (1965) and Bailey and Dell (1973) for maximum likelihood methods, by Zarnoch and Dell (1985) and Dubey (1967) for percentile methods. Other methods were found based on tables (Mann, 1967) and the sample coefficient of variation (Newby, 1980). Neither of the latter methods were shown to be a superior method for parameter estimation.

Bailey and Dell (1973) noted that maximum likelihood estimators are generally the best but for the Weibull distribution the method requires iterative computations. They stressed that the easiest estimators to compute are based on percentiles.

Dubey (1967) and Zarnoch and Dell (1985) presented methods for estimating the shape parameter using two percentile points: 0.17 and 0.97. These two percentiles were found by Dubey (1967) to have 82% asymptotic efficiency, the highest efficiency of the percentiles examined. For estimating the scale parameter b, Zarnoch and Dell (1985) suggested the 63rd quantile minus the estimate of the location parameter (a). For estimating the location parameter (a), they suggested setting it equal to the smallest observed diameter. Dubey (1967), on the other hand, suggested a method for estimating a based on three percentiles. He wrote that his method yielded a value of the location parameter that would not exceed the smallest observation nor would it fall below some theoretical limit.

Cohen (1965) developed the iterative form of the maximum likelihood estimators for the Weibull distribution. The likelihood function of the Weibull pdf was manipulated to form two equations. In one, the shape parameter (c) appeared on both sides of the equation necessitating an initial estimate of c and an iterative solution. When the two sides of the equation converged to equality, the estimated value of c ($\hat{c}$) that brought about the equality was the maximum likelihood estimate of the shape parameter. In the second equation the estimate of c was used to estimate the scale parameter (b). He offered no method for estimating the location parameter (a) in conjunction with b and c. Cohen (1965) only dealt with the two-parameter Weibull pdf which does not include a location parameter. It seemed possible that a likelihood function could be derived for the three- parameter Weibull pdf and that a, b, and c could be estimated more or less simultaneously.

Zarnoch and Dell (1985) presented maximum likelihood estimation as a difficult and costly method. However, Cohen (1965) pointed out that solving for the shape parameter (c) by using the iterative formula could be done effectively using trial and error, giving a simpler and cheaper method. He wrote that it is only necessary to obtain two estimates of c that bracket the final value of $\hat{c}$. The bracketing was explained as the situation when two values of the left hand side of the iterative formula for c are above and below the value of the right hand side. The final step was given as linear interpolation between the two incomplete estimates of c to obtain the actual estimate of c. The vague proviso given was that the two bracketing values of $\hat{c}$ must be within a "sufficiently narrow interval".

The functional forms of these estimation methods were given as follows:

From Cohen (1965) and Zarnoch and Dell (1985) for maximum likelihood estimation, the following methods applied:

1. c is estimated iteratively from

$$[(\sum x_i^{\hat{c}} \ln x_i)/(\sum x_i^{\hat{c}})] - 1/\hat{c} = [(1/n) (\sum \ln x_i)]$$

2. b is estimated from

$$\hat{b} = [(1/n) (\sum x_i^{\hat{c}})]^{1/\hat{c}}$$

From Zarnoch and Dell (1985) for percentile estimation, these methods were given:

1. a is estimated from

$$a = (x_1 x_n - x_2^2)/(x_1 + x_n - 2x^2)$$

where x_i = the i^{th} ordered value (ascending) in the sample, n = the sample size,

x_1 = the smallest ordered value in the sample.

2. b is estimated from

$$\hat{b} = -\hat{a} + x_{.63n}$$

where $x_{.63n}$ = the 63rd quantile of the sample.

3. c is estimated from

$$\hat{c} = \frac{\ln[\ln(1-p_k)/\ln(1-p_i)]}{\ln[(x_{npk} - \hat{a})/(x_{npi} - \hat{a})]}$$

where $p_k = 0.97366$

and $p_i = 0.16731$.

Zarnoch and Dell (1985) investigated the properties of both percentile estimators and maximum likelihood estimators by applying them to artificially generated populations of tree diameters with known Weibull parameters. They found that the magnitude and direction of bias varied according to the parameter for both percentile estimators (PCTE) and maximum likelihood estimators (MLE). They concluded that the maximum likelihood estimators were superior in accuracy to the percentile estimators. They pointed out, however, that the percentile estimators should not be considered unsuitable. They gave as evidence the simple and explicit nature of percentile estimators, and that their behavior when c is near or below 2 is comparable to or better than that of the maximum likelihood estimators. Behavior in this context meant that, although the bias of percentile estimators exceeded that of the maximum likelihood estimators, the asymptotic variances of the percentile estimators were much smaller.

Zarnoch and Dell (1985) concluded further that the Weibull distribution may be insensitive to the magnitude of the bias in percentile estimators. Their results indicated that, although the estimators may have been considerably inaccurate, the population distribution was remarkably well estimated. They noted generally that the maximum likelihood estimators appeared to produce slightly better estimates but the percentile estimators were probably well within the margin of error anticipated by most researchers.

Finally, Zarnoch and Dell (1985) stated that because the parameters of the Weibull distribution are correlated, various combinations of parameters can lead to very similar distributions. They concluded that percentile estimators are entirely appropriate for those interested in the distribution and not in interpreting individual parameters.

B. Johnson's S_B Distribution

Norman L. Johnson introduced three models in 1949 that comprised one coherent system of distributions. His development of these models followed the work of Edgeworth (1898). Edgeworth initiated a method of transformation of variables such that the transformed variables could be considered to have a normal distribution. This method was termed the method of translation. Edgeworth worked only with transformations represented by polynomials. Kapteyn and van Uven (1916), Wicksell (1917), and Rietz (1922) extended the method of translation to include general transformations (Johnson, 1949).

Another approach was taken by Pearson (1895) and Charlier (1905). They each developed systems of curves with the main intent of establishing functions that could produce a wide variety of non-normal distribution curves (Johnson, 1949).

Johnson (1949) carried both of these approaches further to produce three models based on a single general transformation. This system was established to produce a unique curve for any mathematically possible combination of skewness and kurtosis values.

Johnson's rationale for pursuing the method of translation was based on the prominence of the normal distribution (Johnson, 1949). He reasoned that given this prominence and the fact that functions associated with the normal curve are well tabulated, it was natural to try to relate observed distributions to the standard form. He further noted that previously defined translation systems only covered a portion of the range of shapes possible with the well established Pearson (1895) curves.

Johnson (1949) gave the method of translation in general. A function of an observed variable was sought which would be, with sufficient approximation, a normal variable. He noted that normal theory could then be applied to the transformed variable.

Johnson based his transformation on four parameters. The justification for this number of parameters was that having four truly independent parameters would prevent a restricted locus of variation for skewness and kurtosis. The general form of the transformation was given as

$$z = g + d*f((x-e)/l).$$

The parameters g and d were defined as governing the shape of the distribution of x. The parameters w and l were given as the location and scale parameters, respectively.

Three systems resulted from Johnson's work: a three parameter lognormal distribution, a four parameter distribution denoted as S_U, and a four parameter distribution denoted as S_B. The S_U distribution was so termed because the function is unbounded at both ends. The S_B distribution was so termed because its function is completely bounded at both ends.

Hafley and Schreuder (1977) introduced the S_B distribution as a means of modelling tree diameter distributions. They based their choice of the S_B on the fact that it describes a wide range of curve shapes. They also noted that the cumulative percentages of the distribution could be easily obtained without messy integration. Perhaps more importantly, they proposed that the SB may be appropriate for tree diameter distributions because of the bounded nature of the function. The bounded end-points have a logical counterpart in natural diameter limits of trees: tree diameters cannot be less than zero and will not exceed a phenological limit.

Johnson (1949) gave the probability density function of the S_B as

$$f(x) = (d/\sqrt{2\pi})\{1/[(x-w)(w+l-x)]\}\exp\{-1/2[g+d\ \ln((x-w)/(w+l-x))]^2\}$$

based on the transformation

$$z = g + d\ \ln((x-w)/(w+l-x))$$

where d = shape parameter

g = shape parameter

w = location parameter (lower end-point)

l = scale parameter (w + l =upper end-point).

Bowman, Serbin, and Shenton (1981) wrote that very little is known about the S_B distribution. The moments of the distribution were shown by Johnson (1949) to have a very complicated form for which exact solutions are not always possible. Mage (1980) found that when w = 0 the S_B becomes a three parameter distribution bounded between 0 and 1. He stated that this curve is directly analogous to the Pearl-Verhulst logistic growth curve. Beyond this relationship, Bowman, Serbin, and Shenton (1981) and

Johnson (1949) found no general pattern of parameter values which could be said to correspond to other, more familiar distribution curves. They explained that this situation was due to the fact that each combination of parameters produces a curve with unique skewness and kurtosis values.

This latter point led Hafley and Schreuder (1977) to conclude that the S_B occupies a region within the skewness/kurtosis space rather than a point or line. Johnson (1949) and Hafley and Schreuder (1977) showed that the S_B is confined within the skewness/kurtosis graph to a space bounded on one side by the impossible region and on the upper side by the line representing the lognormal distribution (Figure 1). Hafley and Schreuder (1977) used this as evidence that the S_B is theoretically more flexible in shape than the Weibull distribution. They further pointed out that the S_B has four parameters and the Weibull only three. Johnson (1949) mentioned that more parameters can be equated, generally, with greater flexibilty of shape. The S_B's larger number of parameters therefore further suggested that the S_B should generate a wider variety of curve shapes than the Weibull.

Johnson (1949) presented three approaches to parameter estimation for the S_B based on knowledge of the end-points. When both end-points are known, the method of moments was suggested as providing maximum likelihood estimates of the shape parameters. Moments were calculated on the transformed variable. When one, generally the lower, end-point is known, the method suggested was percentile estimation of the other end-point and moment estimation of the shape parameters as when both end-points are known. For neither end-point known the method suggested was percentile estimation by solution of four transformation equations where each z represents a different percentile.

The method for both end-points known was further discussed by Johnson (1949). He stated that this method allowed using the sample data directly to set the values of the location and scale parameters. He suggested that the location parameter be set at the smallest observation and the sum of the location and scale parameters be set at the largest observed value. In his numerical examples Johnson (1949) showed that when dealing with classes where observations are class midpoints, the sum of the scale and location parameters ought to extend to the end of the largest class rather than to its midpoint. This latter point was in reference to obtaining class frequencies using the distribution function.

Johnson gave the transformation of x as

$$f_i = \ln[(x_i - \hat{w})/(\hat{w} + \hat{l} - x_i)]$$

With this transformation he stated that the problem then reduces to that of fitting a normal curve to the observed f_i's. Fitting this curve by moments was shown to produce the maximum likelihood estimates of the shape parameters.

The moments were given as

$$\hat{g} = -\bar{f}/s_f$$

and

$$\hat{d} = 1/s_f$$

where

$$\bar{f} = \sum f_i/n$$

and

$$s_f^2 = \sum(f_i - \bar{f})^2/n.$$

For percentile estimation Slifker and Shapiro (1980) used four symmetrical and equidistant normal variates along with hyperbolic trigonometric functions. This method improved on Johnson's simultaneous equations by providing explicit solutions to all four parameters. Their result was directly descended from Johnson's method for neither end-point known. In addition, they proposed criteria for selecting among the three Johnson models based on relationships between percentiles. They suggested that these criteria would be more suitable than using sample estimates of skewness and kurtosis to determine which curve to fit.

Slifker and Shapiro's (1980) method began with the selection of a value of z>0 of a standard normal variate. Based on this value of z they proposed setting four points as +/-z and +/-3z. For the S_B, the distances between each of the outer and inner end-points would be smaller than the distance between the two inner points. They maintained that this relationship would be the result of the bounding on the S_B.

In detail, Slifker and Shapiro (1980) asked to let the sample quantiles x_{3z}, x_z, x_{-z}, and x_{-3z} correspond to 3z, z, -z, and -3z. Three relationships were then defined as

$$m = x_{3z} - x_z$$

$$n = x_{-z} - x_{-3z}$$

$$p = x_z - x_{-z}.$$

These three measures of distance within the sample distribution were used to establish the following criteria for selection of one of Johnson's models:

if $mn/p^2 > 1$ then fit S_U;

if $mn/p^2 = 1$ then fit S_L, the lognormal;

if $mn/p^2 < 1$ then fit S_B.

Using m, n, and p and hypergeometric trigonometric functions they defined parameter estimators for S_U, S_L, and S_B. For S_B these estimators were given as

$$\hat{d} = z/\{\cosh^{-1}(1/2[(1+p/m)(1+p/n)]^{1/2})\}$$

$$\hat{g} = \hat{d}\sinh^{-1}\frac{((p/n)-(p/m))\{1+p/n)(1+p/m)-4\}^{1/2}}{2((p/n)(p/m)-1)}$$

$$\hat{l} = \frac{p[\{(1+p/n)(1+p/m)-2\}^2-4]^{1/2}}{((p/n)(p/m))-1}$$

$$\hat{w} = \frac{x_z + x_{-z}}{2} - \frac{\hat{l}}{2} + \frac{p((p/n)-(p/m))}{2((p/n)(p/m)-1)}$$

Slifker and Shapiro (1980) based the initial choice of z on sample size. They suggested that for a moderately sized sample z should be set at less than one; as sample size increases, the largest practical choice of z also increases.

For determining actual percentiles from the data, Slifker and Shapiro (1980) showed that percentages P_j are obtained from tabulated z values with j = 3z, z, -z, -3z. The next step given was to obtain the percentile $x^{(i)}$ from $i = nP_j + 1/2$. The final step given was to obtain the value of the observations corresponding to the four observed percentile points and using these to compute m, n, and p.

Mage (1980) followed a somewhat similar line of development but made allowance for the use of four equidistant z values that need not be symmetric. He maintained that the use of four symmetric and equidistant normal variates leads to gross simplification. He also stated that Slifker and Shapiro's (1980) use of hyperbolic trigonometric functions offered more difficulty but was more powerful than Mage's use of natural log functions with z values that need not be symmetrical. In addition, Mage (1980) wrote that the use of symmetric normal variates provides for maximum efficiencies in S_B parameter estimation.

Mage (1980) pointed out a major difficulty in both percentile estimation methods. He wrote that the choice of an initial z value dictates the resulting parameter estimates. He established that different z values yield different parameter estimates. He stressed that the ambiguity of obtaining different parameter estimates by different percentile choices may be unacceptable in some applications. However, from examination of Johnson's (1949) numerical examples as well as those of Slifker and Shapiro (1980) and Mage (1980), it appeared as though a particular range of z values could apply to particular types of populations.

Mage (1980) made a superficial comparison between Johnson's moment estimators, Slifker and Shapiro's percentile estimators, and Mage's percentile estimators. He reached no conclusion but from visual inspection it appeared that there was better agreement between the moment estimators and Slifker and Shapiro's estimators than with Mage's estimators.

Johnson (1949) noted that the S_B was capable of producing two unique curve shapes. He showed that a bimodal or dish shaped curve and a flat topped curve were possible. In general, these curves were taken to be somewhat trivial. However, their existence seemed to further imply the difficulty in using the current, seemingly vague percentile estimation methods for Johnson's S_B.

III. Applications of Distribution Models to Forestry

A. Stand structure

Lorimer and Krug (1983) wrote that the most promising method of indirectly assessing forest age structure has been the interpretation of diameter distributions. They proposed using distribution function parameter estimates and distribution curve shapes as indices for forest age structure.

Lorimer and Krug (1983) found that even-aged stands typically have unimodal diameter distributions. They noted that these stands exhibit varying degrees of positive skewness at a young age but approach a more symmetric distribution with time. All-aged stands were shown to have steeply descending, monotonic diameter distributions that can be represented by the negative exponential distribution function. Multi-aged stands were said to consist of several age classes that may or may not have equal prominence. These stands varied from near normal to irregular negative exponential distributions. The variation depended in part on the proportion of shade tolerant species in the stand.

Shade tolerant species were found by Lorimer and Krug (1983) to have the greatest variety of curve forms, frequently deviating from a symmetric, unimodal shape. The lack of symmetry was attributed to the large number of suppressed trees.

Lorimer and Krug (1983) concluded that for all-aged stands, the Weibull shape parameter (c) is less than 1.0 in most cases and is therefore distinct from the shape parameters of even-aged stands. They decided that it was difficult to distinguish even-aged stands from multi-aged stands through the use of diameter distributions.

Lorimer and Krug (1983) also noted that the diameter distribution of overstory trees alone was near-normal for all species. This was the case even when the total distribution was highly skewed.

Lorimer and Frelich (1984) wrote that the negative exponential distribution has been used with all-aged forest diameter distributions. They stated that a semi-log plot of this distribution produces a straight line which implies an invariant rate of attrition from one size class to the next. This rate was shown to be inappropriate since mortality rates decline precipitously as trees progress from saplings to dominant trees. They provided evidence that it is possible to demand properties of a distribution model that agree with biological conditions associated with a population to be modelled.

B. Forest growth and yield modelling

Bailey (1980) noted that recent growth and yield models have incorporated techniques for predicting changes in diameter distribution with stand age. He described these models as using an assumed distribution that is fitted to stand data. The models were further described as using estimates of distribution parameters to develop least squares (regression) equations to predict the parameters, and the distributions, from stand age.

Bailey (1980) set as a requirement for these models that a diameter distribution remain in a given family over the projection time. This requirement was easily justified in that it would be very awkward to respecify a distribution function, and use parameter predictions, at any step in the projection. From this point Bailey (1980) set out to show that the Weibull distribution exhibits a key property that provides some tentative biological justification for its use.

It was shown by Bailey (1980) that the Weibull distribution allows the assumption of a nonconstant relative growth rate in diameter at any two ages. On the other hand, he showed that Johnson's S_B distribution forces the assumption of constant relative growth rate in diameter at two ages. He concluded, based on his data, that relative diameter growth rate cannot be assumed to be constant over all ages for a given density.

Hyink and Moser (1983) acknowledged that the diameter distribution method for predicting yields and stand structure in even-aged forests was firmly established. They went on to present two approaches to the use of diameter distribution models in growth and yield modelling. One was the parameter prediction model, defined as the process of predicting the future values of distribution function parameters. This step was given to lead to computation of stand average attributes such as volume, basal area per acre, trees per acre, and quadratic mean diameter. The other method was the parameter recovery model, defined as the process of predicting future values of stand attributes and then computing the distribution parameter estimates of the underlying diameter distribution. In general, Hyink and Moser (1983) stated that the parameter prediction models are viewed as being somewhat more informative than the parameter recovery models. Their reasoning was that parameter prediction models allow computation of total stand attribute values as well as their distribution by diameter class.

Hyink and Moser (1983) concluded with a remark that tends to sum up the state of knowledge of distribution models and their use in forestry. They stated that there is poor, if any, understanding of the biological relationships between specific distribution function parameters, the forest populations they characterize, and the characteristics of the site upon which they reside.

C. Other applications

Quang and Burkhardt (1984) presented a method for modelling irregular diameter distribution curves. They focused on the Weibull distribution but showed the general applicability of the method to distribution functions. In essence, they used the cumulative probability function of a distribution and different sets of parameter estimates to model segments of a sample distribution. The segment cumulative functions were then joined to create an overall model of the irregular curve.

Ek, Issos, and Bailey (1975) discussed how to estimate Weibull distribution parameters so as to obtain a particular result beyond that of fitting an observed distribution. They showed that a distribution could be modelled in such a way as to have the model produce a specific quadratic mean diameter. They defined the expected value of the quadratic mean as a quadratic equation involving gamma functions. The distribution parameters were defined as the positive roots of the quadratic equation. No explicit solution of parameters was found possible when the shape parameter (c) is unknown.

Little (1983) has provided the only research to date on fitting a distribution model to tree diameter distributions in a mixed species forest. She developed parameter prediction equations based on stand attributes but pointed out that individual species distributions were still needed to complete the work. She found, however, that the Weibull distribution fit the observed diameter distribution of the mixed stands quite well. The major implication of her work was that the use of distribution models need not be restricted to even-aged, single-species forests or stands.

Zarnoch, et al. (1980) used the Weibull distribution to model changes in red pine diameter distributions under different thinning treatments. They looked at proportions of trees per acre and basal area per acre estimated by the Weibull function for the various thinnings. They found that the Weibull provided an adequate fit to observed diameter distributions.

Materials and Methods

I. The Sampling Method

During the summers of 1983-1985, eighty forest stands were selected and sampled as part of the Ecological Classification System (ECS) study - a cooperative agreement between Michigan State University and the U.S. Forest Service. The 80 stands are composed of upland hardwoods and many are in "late successional" stages. The sampling was done on the western unit of the Huron-Manistee National Forest located in the northwestern portion of the Lower Penninsula of Michigan. This area includes parts of Newaygo, Lake, Wexford, and Manistee counties.

Stands were selected at random from a list of stands which exhibited the appropriate overstory characteristics. The list was compiled by air photo interpretation and ground reconnaisance. The critical overstory characteristics for the ECS study were defined as composition, degree of disturbance, age, basal area, and degree of aspen presence.

The overstory was limited to well-stocked upland hardwoods of at least forty years of age. The aspen component was restricted to 20 % of total basal area or less. Disturbance was defined as evidence of harvesting (or other cutting) or fire in the last forty years. Evidence of grazing, insect attack or disease were also cause for rejecting a stand. Stand size was defined as a minimum of 2.5 acres. Other criteria for stand selection were that the stand have some consistency of slope and aspect and that it be on Forest Service land.

Prior to identifying possible sample stands, the sample region was stratified by landform. Stands were then located randomly within the landform strata. A minimum number of sample stands was identified for the strata and for overstory community types. These types were defined as upland oak, mixed oak-red maple, and northern hardwoods.

Once a stand was located on the ground and selected, simple random sampling was used. Permanent markers were placed at point center of all sample points as well as at a reference point that was described in stand summary notes. Four sample points were used in ECS stands 8-80 (stands numbered consecutively by chronology) while 6 sample points were used in stands 1-7. The initially larger sample size was used to assess the level of variability within stands. Results of this assessment allowed reduction of the sample size.

The main sample point was randomly located from the reference point. The three or five remaining points were located as satellites at random azimuths and distances from the main point. Sample points were rejected if they fell outside the stand boundary. In such cases, a new random azimuth and distance was chosen and followed from the main point.

The overstory data which were collected included tree species, diameter, height, age, ten-year diameter increment, crown ratio, and crown class. Understory species abundance and cover were also sampled as part of the inventory. Additionally, soil profile was described and textural samples taken for laboratory analysis.

At each sample point, variable radius plot sampling (point sampling) was performed. A basal area factor (BAF) of 10 (English) was used on all sample points in all stands except stands 60-66. Stands 61-66 were sampled using a BAF of 5 and stand 60 was sampled with a BAF of 5 for two points and a BAF of 10 for two points. The variable of interest in this study, diameter at breast height (dbh), was measured to 1/10 inch on all tally trees at a sample point. The minimum dbh was 3.5 inches, the lower limit of tree merchantability. Data was recorded on U.S. Forest Service-style tally sheets, one sheet per point.

The intent of stratifying and selecting stands as described was to eventually segregate stands based on landform overstory community type, ground flora species cover and abundance, and soil characteristics. Via this post-stratification, stands were grouped and their overstory productivity levels determined from per acre stand averages. The goal of this process was to identify strata characteristics by which land units could be identified with regard to their potential productivity. The potentiality was derived from the "late successional" nature of many of the stands sampled.

The groups of stands were termed Ecological Land Type Phases (ELTP). Of the eighty stands sampled, 72 were classed as one of eleven ELTP's. The remaining 8 stands were not classified due to irregularities in overstory composition and/or other characteristics that did not conform to ELTP definitions. The current ELTP classifications (Table 1) were not given as final but only as an initial step in the process of defining the classification scheme and system. These unofficial classifications were issued in March of 1986. Additional samples were expected to be added to the ECS data base and these stands and further investigation could alter the current ELTP's. This study utilized these current ELTP's to establish an initial basis for ELTP diameter distributions and a methodology by which distribution modelling may be accomplished as ELTP's develop.

Each ELTP was given a numeric code. The first digit of the code indicated the potential late successional overstory community. The first number was also strongly related to soil development within sandy soils.

ELTP 1: pin oak - white oak

ELTP 1_: black oak - white oak

ELTP 2_: mixed oak - red maple

Table 1 : ECS Stand Assignments by ELTP

ELTP	ECS STANDS
1:	48, 49, 50, 53, 54, 55, 65, 78, 79
10:	1, 3, 28, 33, 38, 61, 63
12:	29, 30, 34, 51, 59, 66, 71, 75
20:	15, 18, 39, 45, 46, 76
21:	8, 11, 47, 52, 58, 67, 68, 70, 80
35:	14, 16, 17, 20, 60, 64, 69, 74, 77
37:	4, 5, 10, 44, 72
40:	7, 36, 40, 41, 43, 56, 57
43:	21, 23, 25, 26, 31, 35
45:	6, 22, 24, 27, 32, 37
47:	2
No ELTP:	9, 12, 13, 19, 42, 62, 73

ELTP 3_: red oak - red maple

ELTP 4_: northern hardwoods

The second digit in the code described the soil substrata characteristics relative to plant associations.

0: no textural substrata

1: bands of sandy loam or coarser material

2: subirrigation

3: bands of sandy clay loam or finer materials in ELTP's undifferentiated by ground flora

5: sandy clay loam or finer bands beneath ELTP's with diagnostic ground flora.

A brief description of each ELTP can be found in Appendix I.

II. Data Processing

All overstory sample data collected were entered into computer files as well as retained as original and photocopies of original tally sheets. Individual files were created for each stand. Species codes and dbh data were extracted from these files to create stand files in a statistical analysis program. The data in these files remained as species and dbh of sample tally trees.

Stand files of species and dbh were grouped according to ELTP groupings. The ELTP files then contained total tally trees for all stands in an ELTP. The statistical program was used to produce frequency distributions of tally trees by one inch dbh classes. The dbh classes were defined as being centered on the inch with width of plus 0.4 inch and minus 0.5 inch. For example, the 4 inch class had width 3.5-4.4 inches.

The sampling design truncated the diameter samples at 3.5 inches. Samples of trees less than 3.5 inches in diameter, though available, were found to be inappropriate for this study. Therefore, the population of interest was defined as all trees in ECS stands of diameter 3.5 inches and greater.

Once the frequency distributions of tally trees within an ELTP were obtained, the number of tally trees in each diameter class in an ELTP were transformed to per acre values by the function:

tally trees * [{(baf/ba)/# points}/#stands in an ELTP]

where baf = basal area factor

ba = basal area = $0.00545415 * dbh^2$

as derived from Husch, Miller and Beers (1982). This conversion factor resulted in unbiased estimates of trees per acre for point sampling.

All stands with the same BAF over all points and the same number of points were processed as a group to the point of obtaining per acre frequencies by diameter class. There were four such groups: 10 BAF and 4 points, 10 BAF and 6 points, 5 BAF and 4 points, and one stand (60) with two points at 10 BAF and two points at 5 BAF. Frequency distributions for each ELTP were then compiled by combining per acre frequencies from appropriate groups.

A. Point Sample Data and Distribution Modelling

The per acre frequency distributions of tree dbh by ELTP were used to develop modelled dbh distributions using the Weibull distribution model and Johnson's S_B distribution model. This methodology agreed with the theoretical foundation developed by Van Deusen (1986). He used relative frequency - percentage of total trees in a diameter class - to develop the theory behind the use of point sample data to model dbh distributions. That development parallelled the use of absolute frequencies - the actual number of trees in a diameter class - in modelling dbh distributions, as was done in this study.

The contrast in methodology was in using observed tally frequencies as the basis for modelling and then expanding the predicted tally frequencies to per acre values. Van Deusen's point was that point sampling, or sampling with probability proportional to size, has dbh^2 as the variable in the proportional function. The dbh^2 term was applied directly in transforming observed tally frequencies to modelled per acre frequencies. In the contrasting, incorrect methodology, the dbh^2 proportional transformation would have to be applied after the modelling portion of the process. Because the method used here modelled the per acre frequencies, the dbh^2 term was incorporated in the modelling, following exactly the theory developed by Van Deusen (1986).

B. Calculation Tools

The criteria set forth in the objectives of this study included that the calculations be kept relatively simple. The goal was to provide methodology making the modelling of diameter distributions readily accessible, concise, and yet reasonably accurate. To that end, all calculation was done using a simple spreadsheet program and/or a programmable calculator.

A simple spreadsheet program was used in carrying out the preceeding and following methods. The program was considered to be a standard sort of spreadsheet that would be available to most foresters interested in diameter distributions. More complex spreadsheets were noted to be capable of allowing an individual to perform the calculations more quickly and easily. A programmable calculator was also found to a capable tool for use in carrying out the necessary calculations. Obviously, the time required in the use of a calculator was greater than for a spreadsheet.

The key element in the use of the spreadsheet was defining a series of "macros": individual keys or short sequences of keys that stored longer typed formulas or a series of commands to the program. The use of macros eliminated the need to type in long formulas and command sequences repeatedly. The macros were saved to diskette and loaded with the spreadsheet program.

III. Distribution Model Development

A. Weibull Distribution

The Weibull probability density function was given earlier as

$$f(x) = (c/b)\ ((x-a)/b)^{c-1} \exp[-((x-a)/b)^{c}]$$

where x = dbh

a = smallest dbh in distribution

b = 63rd percentile of the distribution

c = shape parameter.

In order to apply this function to modelling diameter distributions, the parameters a, b, and c first had to be estimated. The conceptual approaches to parameter estimation for the Weibull distribution have been presented.

The method for estimation from percentiles was given by Zarnoch and Dell (1985). Estimation of a, the location parameter, was done according to the definition of the population of interest. That is, since the samples were truncated at a diameter of 3.5 inches, the smallest possible diameter, $\hat{a}$ was set at 3.5 inches. This value then became the estimate of a.

Estimation of b, the scale parameter, was given by Zarnoch and Dell (1985) as

$$\hat{b} = -\hat{a} + x_{.63n}$$

where n = total number of trees per acre

$\hat{a}$ = the estimate of a, the location parameter

$x_{.63n}$ = the 63rd percentile of the distribution.

They also defined the estimator of c, the shape parameter, as

$$\hat{c} = \{\ln\,[\ln(1-p_k)/\ln(1-p_i)]\}/\{\ln[(x_{npk}-\hat{a})/(x_{npi}-\hat{a})]\}$$

where p_k = 0.97366

p_i = 0.16733.

Cohen (1965) and Zarnoch and Dell (1985) gave the maximum likelihood estimators of b and c as

$$[(\sum x_i^{\hat{c}} \ln x_i)/(\sum x_i^{\hat{c}})] - 1/\hat{c}$$

$$= [(1/n)(\sum \ln x_i)]$$

from which c is derived through iteration, and

$$\hat{b} = [(1/n)(\sum x_i^{\hat{c}})]^{1/\hat{c}}$$

In both formulas $x_i = x_i - \hat{a}$.

Percentile estimation was given by Cohen (1965) and Zarnoch and Dell (1985) as a good way to obtain an initial value of $\hat{c}$ upon which to base the iterative solutions of the first equation for maximum likelihood estimation. They noted that when the two sides of the equation are equal, the value of $\hat{c}$ used to obtain this equality is the maximum likelihood estimate of c.

For both estimation methods, grouped frequency counts were the basis for determining parameter estimates. Though this approach ran the risk of suppressing possibly important distributional information as would be available in a complete data list, the simpler nature of this approach was in accordance with the objectives of the study.

For percentile estimation a list of observed cumulative per-acre frequencies and corresponding diameters were obtained for each ELTP. Total number of trees per acre (n) was then multiplied by the two values of p given by Zarnoch and Dell (1985). This provided two cumulative frequency values, x_{npk} and x_{npi}. The diameter classes corresponding to those two cumulative frequencies were then used to calculate c. The same n for the ELTP was then multiplied by 0.63 to gain one more cumulative frequency. The corresponding diameter class for this value was then used to estimate b. All calculations were easily done on a programmable calculator.

For maximum likelihood estimation, diameter class and frequency data for an ELTP were entered into a spreadsheet. Diameter class was necessarily transformed by subtracting 3.5 from all classes. Three other values (columns) were calculated: frequency * ln x, frequency * $x^{\hat{c}}$, and frequency * ln x * $x^{\hat{c}}$. Columns were summed to obtain the elements of the maximum likelihood estimation formula.

The percentile estimate of c was obtained first. That estimate of c was used to solve for the maximum likelihood estimate of c through the iterative formula. In most cases it was necessary to complete only five iterations. At that point values for each side of the equation were sufficiently close (identical in value up to the fourth significant digit) that linear interpolation, as suggested by Cohen (1965), yielded the correct result. The estimate of b was calculated in the same spreadsheet.

The parameter estimates from each method for each ELTP were then used to calculate f(x) for the Weibull distribution. Number of trees per acre by diameter class (absolute frequencies) were obtained by multiplying f(x) by the total observed number of trees per acre. Tables of observed and predicted relative and absolute frequencies were compiled for each ELTP. Chi-squared and Kolmogorov-Smirnoff one sample goodness of fit test statistics were computed for each parameter estimation method in each ELTP. In addition, ELTP level attributes were calculated to assess the predictive power of the parameter estimation methods and the model beyond simple curve fitting.

B. Johnson's S_B Distribution

The S_B distribution probability density function was given earlier as

$$f(x) =$$

$$(d/\sqrt{2\pi})\{1/[(x-w)(w+l-x)]\}\exp\{-1/2[g+d\ \ln((x-w)/(w+l-x))]^2\}$$

where x = dbh

d = shape parameter

g = shape parameter

l = scale parameter

w = location parameter (smallest diameter).

The parameters to be estimated were discussed earlier.

The method for percentile estimation was given by Slifker and Shapiro (1980). Estimation of the parameters was noted as being dependent upon the selection of an initial z-value, where z is a unit normal variate as given in tabular form by Steele and Torrey (1980). Sliker and Shapiro (1980) let this initial z be the basis of four symmetrical values of z : +/-z and +/-3z. The probabilities associated with each of these values, designated P_j, were then used to determine the distribution percentile $x^{(i)}$ from $i = nP_j + 1/2$, where n equals the total number of trees per acre.

Using the four percentiles corresponding to the four z-values, Slifker and Shapiro (1980) developed three relationships

$$m = x_{3z} - x_z$$

$$n = x_{-z} - x_{-3z}$$

$$p = x_z - x_{-z}.$$

These three relationships were used in the hyperbolic trigonometric functions to estimate the four parameters:

$$\hat{d} = z/\{\cosh^{-1}(.5[(1+p/m)(1+p/n)]^{1/2})\}$$

$$\hat{g} = \hat{d}\,\sinh^{-1}\frac{((p/n)-(p/m))\{(1+p/m)(1+p/n)-4\}^{1/2}}{2((p/n)(p/m)-1)}$$

$$\hat{l} = \frac{p\{[(1+p/n)(1+p/m)-2]^2-4\}^{1/2}}{((p/n)(p/m))-1}$$

$$\hat{w} = \frac{x_z + x_{-z}}{2} - \frac{\hat{l}}{2} + \frac{p((p/n)-(p/m))}{2((p/n)(p/m)-1)}$$

Slifker and Shapiro (1980) suggested an initial z-value less than one. In the process of using this method of parameter estimation, a set of tables were developed giving a range of z-values from 0.50 to 0.90 in increments of 0.10 (Appendix II). The tables included sample size and z-value, and gave the four appropriate cumulative frequencies for the four corresponding z-values. Parameter estimation was then simply a matter of choosing a z-value within 0.50 through 0.90 for a particular ELTP, obtaining values for m, n, and p, and calculating the estimates.

The method for maximum likelihood estimation was given by Johnson (1949). It was based on the distribution transformation function and so reduces the problem to fitting a normal distribution. Johnson (1949) defined the transformation function as

$$f_i = \ln[(x_i - \hat{w})/(\hat{w} + \hat{l} - x_i)],$$

the moments of which yielded estimates of d and g, the shape parameters. The moments were given as

$$\bar{f} = \sum f_i/n$$

and

$$s_f^2 = \sum (f_i - \bar{f})^2/n$$

and the estimators as

$$\hat{g} = -\bar{f}/s_f$$

and

$$\hat{d} = 1/s_f.$$

This method required that values for $\hat{w}$ and $\hat{l}$, the location and scale parameters, be estimated or known prior to estimating d and g. As in the case of the Weibull distribution, the most straight forward approach was to define the range of diameters given by the samples as the population range of diameters. Therefore, the location parameter, w, was set at 3.5 and the scale parameter, l, was set at the maximum observed diameter class for an ELTP minus 3.5. Because w+l defines the upper end of the S_B distribution, this method of setting $\hat{w}$ and $\hat{l}$ gave the observed range. In addition, because the S_B distribution has high contact at both ends of its curve, the upper distribution end was defined as the upper bound of the maximum diameter class plus 0.5.

In the case of each parameter estimation method, the grouped frequencies of the observed distributions were used in computation. Again, some specificity in the complete data list may have been lost due to this approach but the conciseness of calculation and data handling was considered to be an important aspect of the study.

Maximum likelihood parameter estimation for the S_B distribution was carried out with the use of a spreadsheet. Diameter classes and frequencies were entered, the transformation f_i was calculated for each class and summed, the moments were obtained from these sums, and the parameter estimates calculated. Parameter estimates from each method were used to calculate f(x) for each ELTP. Absolute frequencies were obtained by multiplying f(x) by total observed trees per acre for an ELTP. As in the case of the Weibull, goodness of fit statistics and ELTP level attributes were calculated to assess the quality of prediction for the model and each parameter estimation method.

IV. Data Analysis

A. Goodness-of-Fit Statistics

Two tests were used to statistically determine the goodness-of-fit of the predicted diameter distributions to the observed distributions. The first was the Chi-squared test and the second was the Kolmogorov-Smirnoff one sample test.

Steele and Torrie (1980) and Conover (1980) defined the Chi-squared test statistic as

$$X^2 = \sum((\text{Observed} - \text{Expected})^2)/\text{Expected}$$

with degrees of freedom equal to the number of classes minus one, minus the number of parameters estimated from the data. They specified that the test statistic be compared to tabulated values of Chi-squared based on a predetermined alpha level and degrees of freedom.

Steele and Torrie (1980) noted difficulties with the test statistic when there are class frequencies less than one. They suggested that consecutive classes with frequencies less than one be combined in both observed and expected distributions. However, Steele and Torrie (1980) and Conover (1980) agreed that there is no accepted protocol for handling frequencies less than one.

For testing the predicted or "expected" diameter distributions against the observed ELTP distributions, consecutive classes with frequencies less than one were combined. The Chi-squared test statistic was then calculated. Degrees of freedom were determined by the number of classes after combining. Two degrees of freedom were subtracted for parameters estimated for the Weibull distributions, both parameter estimation methods, and for the S_B distribution, maximum likelihood estimation. Four degrees of freedom were subtracted for the S_B distribution obtained through percentile estimation.

Alpha level was set _a priori_ at 0.05. The precedent found in the literature for diameter distribution modelling was 0.10. Almost all such studies were conducted on even-aged, single species forests using fractional area sampling. The stands in this study were mixed species, mixed age stands sampled using probability proportional to size. Under these conditions, it was decided that acceptance or rejection criteria of the goodness-of-fit hypotheses could be relaxed somewhat. By decreasing the value of alpha, the rejection region for the hypothesis test was made smaller thereby decreasing the possibility of incorrect rejection of the null hypothesis. That is, the chance of a Type I error was thereby reduced.

The second test used, the Kolmogorov-Smirnoff one sample test, was discussed by Conover (1980). He defined the test statistic as

$$D = \sup_{x} \mid F_n(x) - F_o(x) \mid$$

where $F_n(x)$ = cumulative relative frequency of the predicted distribution and $F_o(x)$ = cumulative relative frequency of the observed distribution.

The test was presented as one which is concerned with the cumulative distributions: observed and expected. Conover (1980) described it as looking at the absolute value of the differences between the cumulative relative frequencies of the observed and predicted distributions. The largest of these differences was given as the value of the test statistic. Tabulated values of D were referred to where comparison is based on alpha level and sample size. Sample size in this case was defined as the total trees per acre for an ELTP. The large sample approximation, for n>40, was given as $D = 1.36/(n + \sqrt{n/10})^{1/2}$.

The Chi-square test gave a class by class assessment of goodness of fit. The Kolmogorov-Smirnoff test was used to provide secondary goodness of fit information. The Kolmogorov-Smirnoff test was rather sensitive to distributional differences in the smaller diameter classes but rather insensitive to differences in the upper tails. Conversely, the Chi-Squared test was overly sensitive to differences in the upper tails and, in one case, gave an unreliable indication of goodness of fit. The Kolmogorov-Smirnoff test was used to provide further goodness of fit discrimination when the Chi-squared test did not provide reliable results.

The Chi-squared test was also applied to that portion of each distribution greater than or equal to 11 inches in diameter. Again, classes with frequencies less than one were combined. The trees greater than or equal to 11 inches in diameter were defined as sawtimber size. Because of size and potential for producing a higher value product than smaller trees, the sawtimber portion of the distributions was considered critical in modelling. The Chi-square test on this portion of the distributions was intended to provide further evidence on the quality of the performance of each model and parameter estimation method.

B. ELTP-Level Parameters

Bailey (1980), Hyink and Moser (1983), and Little (1983) made the point that stand level attributes are of real interest in growth and yield modelling. These attributes included basal area per acre (BA/a), number of trees per acre (not/a), arithmetic mean diameter at breast height (amdbh), and quadratic mean diameter at breast height (qmdbh). The ability of a distribution model to accurately predict frequencies of trees per acre by diameter class was said to be of basic importance to growth and yield modelling. The further ability of a model to predict a distribution that also yields accurate predictions of stand level attributes that are based on easily obtained measurements and are closely correlated with volume was noted to be of equal significance.

Therefore, the second level of model and parameter estimation method assessment was to determine how well the predicted distributions agreed with the observed distributions on ELTP level attributes.

Not/a was calculated by summing the observed and predicted per acre frequencies respectively for each model and parameter estimation method in each ELTP. BA/a was calculated by obtaining the basal area of each diameter class, multiplying by class per acre frequencies, and summing. Qmdbh was calculated by dividing the BA/a by not/a, then dividing that quotient by 0.00545415 and taking the square root of the result. Amdbh was calculated by multiplying diameter class by frequency, summing the products, and dividing by not/a.

No conclusive means of assessing attribute prediction for an individual ELTP was found. Instead, predictions were assessed by model and parameter estimation method through the use of ATEST. Rauscher (1986) developed ATEST, a computer program written in BASIC, to determine the bias of a predictor based on observed values. This bias was described as the difference between observed and predicted values given in units of measure and as a percentage of the observed. The program was described as giving a value for bias of a prediction based on normally distributed differences or non-normally distributed differences. Normally distributed differences were given to lead to use of Student's t in order to establish a confidence interval about the bias. Non-normally distributed differences were given to lead to the use of a trimmed mean and a jackknifed estimate of variance to obtain a confidence interval about the bias. An alpha level of 0.05 was used in constructing the confidence intervals.

Using this program meant that bias and accuracy would be reported for a model and parameter estimation method over all ELTP's. Therefore, attribute prediction was assessed on an overall level. The main interest in using ATEST was to establish whether the prediction was significantly biased and to what degree it was accurate. Whether or not the 95 % confidence interval about the bias contained zero was the criterion for determining if the bias was significantly different from zero.

C. Skewness and Kurtosis

The final level of analysis was the comparison of observed and predicted values of skewness and kurtosis. ATEST was also used in this comparison.

Steele and Torrie (1980) defined the coefficients of skewness and kurtosis as incorporating the second, third, and fourth moments of the mean as calculated from sample data. These moments were defined as

$$m_2 = n^{-1} \quad (x_i - \bar{x})^2$$

$$m_3 = n^{-1} \quad (x_i - \bar{x})^3$$

$$m4 = n^{-1} \quad (x_i - \bar{x})^4$$

and the coefficients were defined as

$$\text{skewness} = \sqrt{B_1} = m_3/(m_2^{3/2})$$

$$\text{kurtosis} = B2 = m_4/m_2^{2}.$$

They defined skewness as a measure of displacement of the mode of a distribution from centrality. Kurtosis was defined as a measure of the "peakedness" of the distribution.

The three moments were calculated in a spreadsheet by first entering the diameter classes and frequencies for an ELTP. Arithmetic mean diameter was calculated as given earlier. Each diameter was deviated from the mean and raised to the second, third, and fourth power in turn and multiplied by the class frequency. The deviations were then summed and divided by the total trees per acre.

One of the key features of each model was given as flexibility in generating a variety of curve shapes. The SB distribution was said to be more flexible than the Weibull, i.e., that it could generate a wider variety of curve shapes. The intent of this comparison was to bring to light information of an ancillary nature regarding any questions arising from differences in model performance.

Results and Discussion

The results of this study may be put into three categories: model results, goodness-of-fit results, and ELTP-level parameter predictions. Results from the application of the distribution models and parameter estimation methods include stand tables of observed and predicted trees per acre and accompanying distribution graphs. This section also includes discussion of the models and methods used from the standpoint of application mechanics. Results of goodness-of-fit tests include test statistics and outcomes and a discussion of the relative performance of models and methods with respect to observed values. Prediction results include predicted ELTP-level parameter values, relative measures of bias with respect to observed values, and discussion of the significance of the parameters and their prediction through diameter distribution modelling.

I. Modelling Results

Tables 2 and 3 present the results of parameter estimation for the Weibull and S_B distributions, respectively. The non-unique nature of estimates obtained by percentile estimation is apparent in the Weibull distribution. ELTP's 1 and 10 have identical shape parameter estimates as do ELTP's 20 and 35. In addition,

Table 2 : WEIBULL DISTRIBUTION PARAMETER ESTIMATES

ELTP	PERCENTILE ESTIMATES $\hat{a}$	$\hat{b}$	$\hat{c}$	MAXIMUM LIKELIHOOD ESTIMATES $\hat{a}$	$\hat{b}$	$\hat{c}$
1	3.5	3.5	1.535951636	3.5	4.009998	1.397765611
10	3.5	4.5	1.535951636	3.5	4.097647	1.361050734
12	3.5	4.5	0.953222577	3.5	4.116002	1.230833299
20	3.5	5.5	1.409647351	3.5	5.249526	1.425646998
21	3.5	4.5	1.467352544	3.5	4.187545	1.27566721
35	3.5	5.5	1.409647351	3.5	5.226964	1.332776906
37	3.5	7.5	1.317426526	3.5	6.965209	1.395639944
40	3.5	4.5	0.981705321	3.5	4.104463	1.301710538
43	3.5	4.5	0.887603593	3.5	4.675888	1.108687
45	3.5	5.5	1.360272367	3.5	5.232975	1.325986

Table 3 : S_B DISTRIBUTION PARAMETER ESTIMATES

ELTP	PERCENTILE ESTIMATES $\hat{d}$	$\hat{g}$	$\hat{l}$	$\hat{\omega}$	MAXIMUM LIKELIHOOD ESTIMATES $\hat{d}$	$\hat{g}$	$\hat{l}$	$\hat{\omega}$
1	1.208136	3.924946	68.934752	3.532624	0.978188	1.702728	19.5	3.5
10	1.077897	1.847487	22.469508	3.000000	0.908183	1.526177	18.5	3.5
12	0.805918	1.362328	18.526977	3.669845	0.846743	1.503254	19.5	3.5
20	0.876836	1.073885	19.880300	3.334849	0.846743	1.194924	19.5	3.5
21	0.975416	2.413091	32.863353	3.568323	0.914657	1.748078	22.5	3.5
35	0.813321	1.323146	24.105260	3.694737	0.884823	1.493494	23.5	3.5
37	0.811173	1.073024	25.544830	3.004643	0.774972	1.119449	25.5	3.5
40	0.814000	1.357500	17.635000	3.549167	0.853048	1.386943	17.5	3.5
43	0.723467	1.411248	26.033420	3.522514	0.769000	1.496538	24.5	3.5
45	1.216796	2.613756	51.380930	2.309535	0.943734	2.003481	33.5	3.5

ELTP's 20 and 35 have identical scale parameter estimates. On the whole, the Weibull shape parameter estimates show a range of curve shapes from reverse J-shape to a somewhat strongly right skewed mound shape (0.8876 - 1.53595).

The maximum likelihood estimates (MLE) are different from the percentile estimates (PCTE) for both models. The MLE are noticeably more conservative in that they exhibit a narrower range of values. For the Weibull, three of the shape parameter estimates differ in a significant way between estimation methods for the same ELTP. The PCTE give values of c less than 1.0 for ELTP's 12, 40, and 43 whereas corresponding MLE are greater than 1.0. There is a basic change in curve shape as c differs from 1.0; when c is less than 1.0 the curve becomes a reverse J-shaped distribution; when c is greater than 1.0 the curve becomes mound-shaped. For ELTP 12 the PCTE does not correspond to the class of curve of the observed distribution (Figure 6). The reverse is true for ELTP's 40 and 43 (Figures 16 and 18).

Far less can be discerned by examining the parameter estimates for the S_B distribution (Table 3). The main observation that can be offered is that relative magnitudes of parameter estimates are similar between parameter estimation methods. This at least suggests some degree of internal consistency in the model.

The S_B percentile parameter estimation process appears to offer more difficulty in practice than in theory. Initial estimates for ELTP's 1, 10, 12, 20, 21, 43, and 45 are those given in Table 3. These estimates are the direct result of the estimation method described. The estimates for ELTP's 35, 37, and 40 in Table 3 are the result of some modification. Estimates obtained directly by the given method do not produce a distribution of sufficient range in application to this data set. The scale parameter in particular is too small.

This problem is corrected by rejecting the underestimates of the scale parameter. The corrected estimates are values of the scale and location parameters that produce frequencies in a range of diameters identical to that of the appropriate ELTP. For this problem, the sum of $\hat{w}$ and $\hat{l}$ is set to equal the maximum diameter class and the solutions for $\hat{w}$ and $\hat{l}$ are obtained by two or three iterations of f(x), when the proper range is obtained.

The correction of percentile estimates is permissible since percentile estimation is a non-deterministic method. Percentile estimation of parameters is simply a systematic and repeatable method for fitting a curve to an observed frequency curve. For curve fitting only, one may obtain parameter estimates by trial and error alone. The appropriate estimates are those that provide the best curve fit.

Correcting parameter estimates creates two problems. First, the process is time consuming and ill-defined. It is probable that the use of a complete data list instead of grouped frequencies would provide sufficient distributional detail to overcome the problem. Second, when the scale and location parameter estimates are corrected, the shape of the curve is altered. In effect, the curve becomes less kurtotic as the location parameter is decreased and the scale parameter is increased. The fit to the complete observed distribution may then be better or worse. If the shape parameter(s) need to be corrected, estimate correction becomes increasingly complex.

The method of maximum likelihood estimation for the S_B is straight forward, as given. The use of the moments of the transformed variable (diameter) is simple and precise. The only difficulty with this method is that $\hat{w}$ and $\hat{l}$ must be taken from the sample data directly. Since the S_B curve approaches zero very quickly at the bounds the values of $\hat{w}$ and $\hat{l}$ must be set so as to allow sufficient frequencies to occur near the bounds. The method described is adequate for accommodating this feature.

Both estimation methods for the Weibull distribution are easy to apply as described. The MLE method requires somewhat more time than the PCTE method and about the same as the S_B MLE method. Weibull PCTE are the result of about three hours of work, including the production of the frequency tables. The Weibull MLE method currently requires about six hours including obtaining the distribution frequencies. Time savings for this method are possible because of the piecemeal setup of calculations, the limited nature of the spreadsheet used, and the speed limitations on the computer used. The S_B MLE are the result of approximately four hours of work including obtaining frequencies. The S_B PCTE require about six to seven hours of work for the initial estimates. The correction of estimates adds an undetermined amount of time to the process.

The Weibull PCTE method is clearly quicker and is also easier than the other methods. The S_B MLE method is easier and quicker than the Weibull MLE method because the latter is done iteratively. The use of spreadsheet and calculator makes obtaining parameter estimates rather easy, quick, and immediate in the sense that no mainframe computer time is required.

Tables 11 - 20 and accompanying Figures 2 - 21 detail the observed and predicted distributions and their curves. Parameter estimation method results are paired in the graphs. Pairing methods results is preferred because of the inherent differences in the methods. Both relative frequency (RF) and absolute frequency (AF) are given in the tables for each ELTP.

Some aspects of curve behavior can now be linked to parameter estimate values. Although ELTP's 1 and 10 (Tables 11 and 12; Figures 2 and 4) have the same shape parameter estimate via percentile estimation, the difference in their scale parameter estimates causes rather different curves. This difference gives an idea of what effect changing the scale parameter has on the distribution.

The disagreement between percentile and maximum likelihood estimates of the Weibull shape parameter for ELTP's 12, 40, and 43 can be seen in Tables 13, 18, and 19 and Figures 6, 7, 16, 17, 18, and 19. The discrepancy is in fact small and may be related to the truncation of the distribution. Because of this truncation it cannot necessarily be said that one or the other estimate is incorrect. The behavior of the distribution around c = 1.0 is interesting in spite of the truncation. It can be seen that small changes in c around 1.0 definately change the general class of curve that results. At c equal to about one the distribution develops a shoulder on the left tail. When c becomes even slightly less than one the distribution becomes a definite reverse J-shaped curve (Tables 13 and 19, Figures 6 and 19).

In general, visual inspection reveals a reasonable approximation of the observed distributions by the predicted in all cases except ELTP's 35 and 37. These ELTP's will be discussed later.

Closer inspection reveals that the left tail of the observed curves can be highly variable. This variability makes fitting a curve to the observations overall somewhat difficult. Where the observed curves are relatively regular, the predicted curves can be seen to fit much better (Figures 2 - 21). The variability of frequencies in the lower diameter classes is a product of sample size, sampling method, and the stands sampled. A larger sample size would provide greater opportunity for observing smaller diameter trees. The expansion of point sample frequencies to per acre frequencies results in small diameter trees being weighted much more heavily than larger diameter trees. The outcome is that the absence or presence in a sample of a single 4 inch tree greatly affects the expanded frequencies. Finally, in many stands, the inconsistent presence of multiple stemmed red maples of small diameter (< 7") ends up creating a sawtoothed left tail in ELTP's 35 and greater.

Though trees less than 3.5 inches in diameter are not included in the observed frequencies, there is no indication that this truncation presents any problems in modelling. The location parameter for each model provides the means by which a lower bound is placed on the distribution. Frequencies beyond that point are then allocated according to the observed frequencies by either parameter estimation method.

II. Goodness-of-Fit Results

The null hypothesis of the Chi-squared goodness-of-fit test states that there is no difference between observed and predicted distributions. Failure to reject the null hypothesis leads to the conclusion that the predicted distribution is as representative of the underlying population distribution as is the observed distribution. The same null hypothesis and conclusion are applicable to the observed and predicted ELTP diameter distributions generated in this study.

The Chi-squared test examines individual class absolute frequencies to assess goodness-of-fit. The Kolmogorov-Smirnoff test works with cumulative relative frequencies of the distributions. The differences between these tests present different aspects of goodness of fit. The Chi-squared test is used in this study as the primary indicator of goodness-of-fit because it shows how well individual class absolute frequencies are predicted. These class frequencies are of interest in assessing the accuracy of a prediction on the dependent variable of the study: frequencies of trees per acre by diameter class. The Kolmogorov-Smirnoff test is used as an indicator of the accuracy of the distribution in general. As such it can be expected to be less discriminating and more related to the distributions as continuous rather than discrete as they are treated here.

A. Percentile Estimation

Table 4 provides the calculated test statistics for the PCTE models. According to the Chi-squared test, both models produced significantly different distributions (alpha = 0.05) for six ELTP's. That only 40% of the observed distributions for each model are non-significant indicates that the PCTE models did not do an adequate job of modelling the observed distributions. This is especially so considering that the alpha level was set to allow greater latitude by reducing the size of the rejection region. However, the hypothesis tests are not the final word on how well the models perform, as is shown in the sections to follow.

The PCTE S_B model produces non-significant results for the so-called low site ELTP's: 1, 10, and 12. The PCTE Weibull model produces no such pattern of non-significance. Both PCTE models produce significant results for the higher site ELTP's - 35, 37, 40, 43, and 45 - with the exception of the Weibull PCTE model for ELTP 45. Over all ELTP's the S_B PCTE model produces somewhat lower Chi-squared scores. Considering that three of the S_B distributions have corrected parameter estimates, the lower scores over all ELTP's is a point of interest. However, because of the increased difficulty in parameter estimation, the small improvement over the Weibull PCTE model over all ELTP's may not be worth noting.

Table 4 : GOODNESS OF FIT STATISTICS FOR PREDICTED DIAMETER DISTRIBUTIONS

	WEIBULL PCTILE				Sb PCTILE			
ELTP	CHI^2	df	K-S	n	CHI^2	df	K-S	n
1	55.80*	11	0.0784	252	15.87	9	0.0966*	252
10	12.53	10	0.0722	261	7.86	8	0.0411	261
12	22.88*	12	0.0859*	261	12.96	10	0.0438	261
20	17.64	12	0.0268	234	20.91*	10	0.0550	234
21	18.64	12	0.0682	305	16.71	10	0.1111*	305
35	32.82*	13	0.0965*	264	36.39*	11	0.0569	264
37	66.73*	14	0.0848	179	72.87*	12	0.1484*	179
40	25.72*	10	0.0832*	311	18.10*	8	0.0819*	311
43	24.33*	13	0.0548	262	25.54*	11	0.0739	262
45	12.22	11	0.0247	278	24.00*	9	0.0902*	278

alpha = 0.05 * indicates significance
K-S = Kolmogorov-Smirnoff test statistic

As a final note on the S_B PCTE Chi-squared scores, the degrees of freedom are two fewer than for the Weibull. The reduction in degrees of freedom increases the size of the rejection region for the goodness-of-fit test from that of the Weibull. The fewer degrees of freedom therefore makes statistical fitting of the S_B PCTE model more difficult than for the Weibull.

Rejection of the null hypothesis in the Chi-squared test is of prime concern. As such, the Weibull and S_B PCTE distributions model the observed distributions about equally well. This result indicates that both models are about equally accurate in predicting individual class frequencies for ELTP's, though the models do not model the same ELTP's equally well.

Rejection of the null hypothesis in the Kolmogorov-Smirnoff test is of concern for it indicates a lack of cumulative distributional accuracy. For this study, this result is considered less important than the results of the Chi-squared test. However, the Kolmogorov-Smirnoff test provides information on the general suitability of the models to this application which can lead to general comments on model selection.

The two tests do not agree in all cases (Table 4). The Kolmogorov-Smirnoff test is more sensitive to differences in the small diameter classes. However, in many cases the cumulative distributions approach each other reasonably well to produce better agreement than indicated by the Chi-squared test. The Komogorov-Smirnoff test indicates that although the two models produce about equal accuracy in class frequency prediction, the Weibull PCTE model provides a better overall fit than the S_B. This is likely due to the corrections made to the S_B parameter estimates in order to obtain a suitable diameter range. The resulting distributions may be less accurate overall than if corrections were not necessary.

The Weibull PCTE model must be preferred over the S_B PCTE model in this study based on goodness-of-fit results. Both goodness-of-fit tests indicate that the Weibull PCTE model performs at least as well as the S_B PCTE model. This is the telling point. The greater difficulty of estimating parameters for the S_B PCTE model requires that it perform better than the Weibull in order for its use to be justified. Since it performs only about as well at best, nothing appears to be gained by the increased complexity.

B. Maximum Likelihood Estimation

The distributions derived through maximum likelihood estimation of parameters do a much better job of fitting the observed data (Table 5). Though this method of parameter estimation is more difficult in general than the percentile estimation methods, the improvement in accuracy justifies its consideration.

Table 5 : GOODNESS OF FIT STATISTICS FOR PREDICTED DIAMETER DISTRIBUTIONS

ELTP	WEIBULL MLE CHI^2	df	K-S	n	Sb MLE CHI^2	df	K-S	n
1	8.79	11	0.0251	252	11.69	11	0.0577	252
10	4.79	10	0.0948	261	10.76	10	0.0741	261
12	11.57	12	0.0405	261	13.35	12	0.0700	261
20	18.14	12	0.0470	234	22.40*	12	0.0827	234
21	6.95	12	0.0251	305	8.33	12	0.0635	305
35	30.39*	13	0.0665	264	31.45*	13	0.0811	264
37	66.46*	14	0.1377*	179	80.89*	14	0.1778*	179
40	11.29*	10	0.0461	311	14.95	10	0.0759	311
43	19.96	13	0.0475	262	23.13*	13	0.0991*	262
45	11.43	11	0.039	278	28.58*	11	0.0937*	278

alph = 0.05   * indicates significance
K-S = Kolmogorov-Smirnoff test statistic

Both the Chi-squared and Kolmogorov-Smirnoff tests show that the MLE models perform decidedly better than the PCTE models. The improvement in performance can be determined by the relative magnitudes of the test statistics. Only in the case of ELTP's 35 and 37 is there no improvement. Both of those distributions are so irregular that curve fitting of any regular sort would prove difficult or impossible. Examination of the distribution data and curves in Tables 11 - 20 and Figures 2 - 21 gives an idea of how close in agreement the MLE models come to most of the observed distributions.

On the whole, the Weibull MLE model does somewhat better than the S_B MLE model. The differences could be explained by noting that two of the S_B parameters were taken directly from the sample data while the Weibull required only one parameter to be treated in that way. The setting of parameter estimates from sample data may introduce additional error into the estimation process. In making a choice between these two models, it should be acknowledged that differences in accuracy are not substantial. Instead, the basis for choice should be that although the iterative calculations for estimating the Weibull shape parameter are somewhat time consuming, that method is more exact than relying on setting sample observations as parameter estimates. Therefore the Weibull MLE model is indicated as superior to the S_B MLE model on the basis of goodness-of-fit tests on the whole distribution.

Goodness-of-fit tests on the whole distribution are, however, not the final word in determining the adequacy of a model. Hafley and Schreuder (1977) use goodness-of-fit criteria as a relative measure of distribution adequacy. Still other studies (Little, 1983; Zarnoch, et al., 1980; Johnson, 1949) do not base their findings completely on goodness-of-fit tests, or also are interested in how well predicted distributions estimate certain aspects of the data. It is typical that distribution modelling is a step in dealing with a larger estimation or modelling problem. It is the nature of the data itself that dictates what is important in modelling. In this case, diameter distributions are presented as possibly leading to forest growth and yield modelling. Other criteria in addition to goodness-of-fit to the entire distribution are of interest for assessing the relative worth of a distribution model for growth and yield modelling. These criteria include how well the model predicts frequencies of larger diameter, higher value trees and how well the model predicts forest level attributes. Both of these criteria are not mutually exclusive of the overall goodness-of-fit tests but need not follow the pattern of accuracy established by the tests.

III. Goodness-of-Fit for Classes 11 Inches and Greater

Trees 11 inches in diameter and greater are defined as sawtimber size and as such demand attention as potentially higher value trees than those less than 11 inches. Table 6 presents Chi-squared goodness-of-fit scores for the two models and two parameter estimation methods. The Kolmogorov-Smirnoff test is not included because the concern is with accuracy of class frequency prediction alone.

Results are much improved over the full distributions. The percentile estimation models produce seven non-significant distributions. Because degrees of freedom serve to scale the rejection region, the results between the full distributions and the sawtimber segments may be compared. Visual inspection of the curves indicates that the sawtimber segments are more regular than the smaller size class segments. The more regular curve would be easier to model on its own. As it is, both models predict the sawtimber segment quite well as a part of a total distribution.

The MLE models are again superior in prediction to the PCTE models. For this segment, there is no real difference between models for a parameter estimation method. Therefore, the choice of model must go to the Weibull since it provides the easier and more exact means for parameter estimation.

Table 6 : GOODNESS OF FIT OF PREDICTED DISTRIBUTIONS FOR DIAMETER CLASSES 11" AND GREATER

ELTP	WEIBULL PCTILE CHI^2	df	Sb PCTILE CHI^2	df	WEIBULL MLE CHI^2	df	Sb MLE CHI^2	df
1	33.98*	4	1.89	2	4.95	4	4.11	4
10	1.04	4	1.13	2	0.49	4	1.49	4
12	11.62	6	5.73	4	4.84	6	5.23	6
20	8.54	6	11.49*	5	9.19	6	10.76	6
21	8.27	6	3.37	4	3.27	6	1.80	6
35	2.17	7	13.90*	5	3.71	7	8.44	7
37	18.38*	8	27.89*	6	14.21	8	30.19*	8
40	12.72*	4	5.40	2	4.23	4	4.04	4
43	9.38	7	9.47	5	5.96	7	6.09	7
45	3.72	5	13.33*	3	3.54	5	12.51*	5

alpha = 0.05 * indicates significance
df = degrees of freedom

ELTP 37 stands out as the most intractable in all cases. The observed distribution is bimodal and reaches its maximum diameter class somewhat suddenly (Table 17 and Figures 14 and 15). In addition, both modes appear quite leptokutotic, making the curve even more unbalanced. The reasons for this are not entirely clear. A breakdown of species distributions reveals that the modes are not especially associated with individual species or particular groups of species. Age data is incomplete so it is unknown whether the modes coincide with different age groups. A breakdown by stands shows that three of the five stands exhibit the identical modes of the ELTP distribution, one at 4 inches and another at 10 inches. The other two stands have no trees in the 4 inch class and have their respective modes at 10 inches. Given the shape of the observed curve, there is almost certainly no way to adequately model it using the methods described in this study.

IV. ELTP-Level Parameters

The final level of assessment looks at prediction of the parameters basal area per acre (ba/a), total trees per acre (not/a), arithmetic mean diameter (amdbh), and quadratic mean diameter (qmdbh). All of these parameters are directly related to the distributions from which they are calculated. However, they are summaries of the distributions and as such can be different in accuracy than the predicted curves from which they are derived.

These parameters are considered important for what their prediction accuracy indicates about the respective models and parameter estimation methods. The goodness-of-fit tests deal with the details of the distributions. Consideration of ELTP-level parameter predictions leads to statements about how well the modelled distribution as an aggregate characterizes an ELTP in a standard manner of summary. The ability to predict and therefore project forest level parameters is considered a key to obtaining the best results from growth and yield modelling (Little, 1983).

The observed and predicted estimates of the ELTP-level parameters are given in Tables 7 and 8. The error in prediction (bias) and the confidence limits about that error are presented in Table 9. Predictions are by ELTP and distribution parameter estimation method by model. Errors and confidence intervals are for ELTP-level parameter predictions over all ELTP's.

By and large, the predictions are adequate in accuracy, the most extreme deviation being for ba/a for the S_B PCTE model. The most extreme value of the confidence interval for that prediction is 17% of the mean of the observed ba/a. Though accuracy is generally good, several of the predictions are slightly to clearly biased, generally upwards. The preponderance of biased predictions

Table 7 : OBSERVED AND PREDICTED AVERAGES FOR ELTP-LEVEL PARAMETERS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
ELTP	MDBH	NOT/a	MDBH	NOT/a	MDBH	NOT/a	MDBH	NOT/a	MDBH	NOT/a
1	7.14	251.78	6.61	255.22	6.87	252.57	7.11	254.72	7.09	257.02
10	7.25	260.89	7.51	263.25	7.21	259.97	7.21	263.73	7.15	267.66
12	7.39	260.87	7.75	254.12	7.43	270.55	7.29	263.08	7.20	269.25
20	8.27	234.23	8.41	235.26	8.72	236.17	8.19	235.67	8.29	238.26
21	7.37	304.56	7.54	307.52	7.03	319.45	7.34	307.66	7.28	312.79
35	8.30	263.86	8.46	265.72	8.82	268.47	8.25	265.84	8.20	268.47
37	9.92	179.49	10.24	179.06	9.82	177.09	9.76	180.02	9.75	182.16
40	7.29	310.95	7.62	302.09	7.17	322.51	7.26	313.85	7.16	320.27
43	8.00	262.30	8.04	255.24	8.25	270.60	7.93	263.01	7.86	270.07
45	8.31	277.71	8.50	279.85	9.04	271.25	8.28	279.99	8.30	282.07

MDBH = arithmetic mean diameter
NOT/a = number of trees per acre

Table 8 : OBSERVED AND PREDICTED AVERAGES FOR ELTP-LEVEL PARAMETERS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
ELTP	QMD	BA/a	QMD	BA/a	QMD	BA/a	QMD	BA/a	QMD	BA/a
1	7.65	80.47	6.94	67.02	7.39	75.12	7.59	80.06	7.63	81.67
10	7.76	86.80	7.98	91.45	7.78	85.88	7.72	86.71	7.74	87.55
12	7.97	90.27	8.69	104.77	8.14	97.78	7.91	89.89	7.88	91.30
20	8.92	101.92	9.11	106.58	9.52	116.81	8.86	100.99	9.01	105.41
21	8.01	106.70	8.05	108.82	7.73	102.16	7.90	104.38	7.96	108.22
35	9.04	117.68	9.19	122.30	9.85	141.97	9.01	117.60	9.06	120.08
37	10.75	113.05	11.35	125.77	11.05	117.89	10.75	113.38	11.01	120.45
40	7.84	104.13	8.46	117.84	7.84	108.11	7.78	103.71	7.77	105.54
43	8.93	114.13	9.35	123.12	9.48	132.55	8.85	112.31	8.88	116.39
45	9.10	125.25	9.29	131.85	10.16	152.63	9.05	125.18	9.31	133.30

QMD = quadratic mean diameter
BA/a = basal area in square feet per acre

Table 9 : BIAS OF PREDICTED ELTP-LEVEL PARAMETERS AND STATISTICS

PARAMETER/	WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
STATISTIC	BIAS	95% CI	BIAS	95% CI	BIAS	95% CI	BIAS	95% CI
AMDBH	-.20	(-.42,.01)	-.12	(-.37,.14)	.05*	(.02,.08)	.10*	(.06,.13)
QMDBH	-.24	(-.54,.05)	-.30	(-.62,.03)	.06*	(.03,.08)	-.03	(-.11,.06)
NOT/a	.49	(-2.96,3.93)	-3.60	(-7.87,.67)	-2.09*	(-2.76,-1.43)	-6.19*	(-7.82,-4.55)
BA/a	-6.01*	(-11.77,-.25)	-9.15*	(-17.44,-.86)	.62*	(.02,1.22)	-2.68*	(-4.59,-.77)
SKEWNESS	-.05	(-.32,.23)	-.24*	(-.37,-.11)	-.11	(-.34,.12)	-.19*	(-.38,-.01)
KURTOSIS	.13	(-.81,1.08)	-.26	(-.89,.32)	-.24	(-1.17,.69)	-.08	(-.79,.62)

* indicates bias is significantly different from zero at alpha = 0.05, n = 10.

occurs for the MLE models. Biased prediction is determined by the fact that the 95% confidence interval does not include zero.

Arithmetic mean diameter (amdbh) and total trees per acre (not/a) are used here as direct summaries of the distributions. Their values are given in Table 7. Inspection of observed and predicted values shows that the MLE distributions are very close in prediction of the observed values. The bias and 95% confidence intervals in Table 9 confirms the better accuracy of the MLE distribution predictions compared to the PCTE model predictions. This result is no surprise since the accuracy in MLE model curve fitting is superior. Yet for their lesser accuracy, the PCTE predictions are unbiased for both models. The Weibull and S_B MLE predictions are biased, however slightly, for both amdbh and not/a.

The sign on the bias indicates the direction of the bias. Where the 95% confidence intervals include zero this sign is of no consequence as the bias is not significantly different from zero (alpha = 0.05). Where the confidence interval does not include zero the sign of the bias is of interest.

Both MLE models underestimate amdbh, indicated by the positive sign on the bias (as given by the ATEST program). Both MLE models overestimate not/a, as indicated by the negative bias. The implication is that the MLE models tend to overestimate the frequencies of trees in the small diameter classes and/or underestimate the frequencies in the larger diameter classes.

The amount of bias in the MLE models predictions is small: no more than 3% of the parameter mean for the extreme value of the confidence interval. However, the result is unexpected in view of the precedents set in the literature. Percentile estimation is generally expected to produce distributions with greater bias than maximum likelihood estimation (Zarnoch and Dell, 1985). The predictions considered here are somewhat removed from distribution estimation so that this expectation may not apply. Yet these ELTP-level parameters are directly related to the distributions as summaries.

Quadratic mean diameter (qmdbh) and basal area per acre (ba/a) are more sophisticated summaries than amdbh and not/a and are easier to relate to volume. Observed and predicted values for these ELTP-level parameters are included in Table 8. The errors in prediction and 95% confidence intervals about the errors are in Table 9.

Only the Weibull MLE predictions are biased for qmdbh. As before, the magnitude of the bias is small: the extreme value of the confidence interval is only 1% of the ELTP mean qmdbh. The S_B MLE prediction of qmdbh is the most accurate over all ELTP's. There is less to indicate this result in the distributions and their curves.

All models produce biased predictions of ba/a for all ELTP's. The PCTE predictions are two to three times less accurate than the least accurate MLE predictions. The errors for these predictions are much higher than for any other predictions. As such, the PCTE predictions overestimate ba/a by as much as 17% of the ELTP mean ba/a. The S_B MLE predictions are also overestimates of ba/a. The Weibull MLE predictions underestimate ba/a. The magnitude of bias for both MLE models is relatively small.

In general, the Weibull MLE predictions of ELTP-level parameters are the most accurate. However, all of those predictions are biased. The S_B MLE predictions are only slightly less accurate than the Weibull MLE predictions, in general. These predictions tend to be biased as well, except for qmdbh. The PCTE predictions are generally less accurate than the MLE predictions. Except for the ba/a predictions, they are all unbiased.

From a statistical standpoint, unbiased results are preferred. Therefore, though the MLE models produce more accurate curves and predictions of ELTP-level attributes, the biasedness of their summary predictions makes their superiority less certain. This result cannot be applied to the models in general. Observations on the biasedness of predictions are limited only to the data used in this study.

V. Skewness and Kurtosis

The coefficients of skewness and kurtosis are normally used to assess the normality of a distribution or curve. They are overall measures of curve shape in two dimensions. In this study, these coefficients are used to assess the relative curve shapes of observed and predicted curves. The observed and predicted coefficients are in Table 10. The bias of the predictions and the 95% confidence intervals about the bias are presented in Table 9.

For skewness predictions, the S_B models produce biased results. The biasedness in both cases is the result of overestimating skewness. This overestimation of skewness is seen in the distributions (Tables 11 - 20) and graphs (Figures 2 - 21) as reflected in the overestimation of frequencies in the small diameter classes. The overestimation of skewness may be the result of an underestimation of the distribution scale parameter. All predictions for kurtosis are unbiased for both models. The most accurate predictor of overall curve shape as described by skewness and kurtosis is the Weibull PCTE distribution. This is somewhat surprising given that the S_B distribution was noted to be the more flexible model. Two reasons may be given for this result. First, the range of curve shapes given by the observed data is not very wide. Other than the two irregular distributions, ELTP's 35 and

Table 10 : OBSERVED AND PREDICTED ELTP KURTOSIS AND SKEWEDNESS

ELTP	OBSERVED SKEWNESS	OBSERVED KURTOSIS	WEIBULL SKEWNESS	PCTILE KURTOSIS	Sb PCTILE SKEWNESS	Sb PCTILE KURTOSIS	WEIBULL SKEWNESS	MLE KURTOSIS	Sb MLE SKEWNESS	Sb MLE KURTOSIS
1	1.6843	6.9319	1.0202	4.2000	1.8666	7.5311	1.1721	4.6197	1.3129	4.7008
10	1.1018	4.3751	0.9988	4.0630	1.2794	4.6279	1.1882	4.5793	1.2670	4.3911
12	1.1767	4.2757	1.3449	4.3898	1.2674	4.2177	1.3280	4.9449	1.3571	4.6098
20	0.8098	3.2156	1.0147	3.8178	0.8963	3.1654	1.0389	3.9675	1.0407	3.5321
21	1.4339	5.3091	1.09995	4.4495	1.8537	7.2208	1.3466	5.2442	1.4687	5.2306
35	0.9443	3.7714	1.1204	4.3664	1.2092	4.0216	1.2117	4.6362	1.2718	4.36998
37	0.36699	3.1496	1.0405	3.7735	0.9729	3.2628	1.0451	3.9500	1.0689	3.4708
40	0.9228	3.4579	1.2283	3.9662	1.2609	4.1869	1.2043	4.4804	1.2263	4.1358
43	1.2921	4.6799	1.5868	5.3751	1.4734	4.7918	1.4902	5.4874	1.5012	5.0145
45	1.5424	7.3727	1.2526	5.0330	1.5599	6.0657	1.3025	5.2283	1.6547	6.1989

37, only two general curve types appear. Second, the difficulties encountered in the percentile estimation of the S_B parameters is certain to have given a less than satisfactory indication of that distribution's potential performance. Allied to this point is the fact that maximum likelihood estimation for the S_B distribution is less explicit than for the Weibull. In being less explicit, the method may be subject to additional error in estimation. The result in both cases could be parameter estimates that are not sufficiently accurate. The parameter estimation methods are at fault for the S_B and that places them in a position inferior to the Weibull distribution in this study.

The implications of accurately predicting skewness and kurtosis may be carried further. The correlation between the Weibull shape parameter estimates (both methods) and the skewness and kurtosis of the predicted curves is -0.80 and -0.39, respectively. The two tailed 95% critical correlation value is -0.44, n = 20. The Weibull shape parameter is clearly correlated with skewness. The association is not outstanding but there is evidence to suggest that c is associated with skewness in some real fashion. Correlations between the S_B shape parameters and predicted skewness and kurtosis exhibit a similar association: d and skewness = 0.53; d and kurtosis = 0.67; g and skewness = 0.82; g and kurtosis = 0.90.

It should be expected that shape parameters and the skewness and kurtosis coefficients would exhibit some significant degree of association. Both sets of values are indices of curve shape. It cannot be said if the association shown here is truly linear. The point is that it should also be expected that a distribution function that is more accurate in predicting the observed coefficients of skewness and kurtosis is based on estimates of the distribution parameters that are closer to the values of parameters for the true distribution. This is a hypothesis that cannot be tested here since the underlying population distributions are unknown. However, when all ELTP's are taken as a set, the Weibull models, with parameters estimated as in this study, exhibit better overall accuracy in prediction than the S_B distribution when compared within parameter estimation methods. Based on the above hypothesis, this better accuracy may be traced to the fact that the Weibull models are more accurate and unbiased in predicting observed measures of curve shape.

Table 11 : ELTP 1: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1191	29.98	.1471	37.03	.1145	28.82	.1442	36.31	.1469	36.99
5	.1949	49.08	.2123	53.45	.2627	66.14	.1830	46.09	.2174	54.75
6	.1688	42.50	.2018	50.81	.2023	50.93	.1723	43.38	.1762	44.36
7	.1735	43.68	.1614	40.65	.1364	34.35	.1444	36.37	.1321	33.26
8	.1108	27.89	.1153	29.04	.0903	22.73	.1127	28.38	.0975	24.55
9	.0863	21.75	.0756	19.01	.0602	16.16	.0835	21.01	.0718	18.07
10	.0540	13.60	.0460	11.58	.0408	10.26	.0592	14.92	.0527	13.28
11	.0186	4.68	.0263	6.62	.0280	7.05	.0406	10.22	.0386	9.73
12	.0191	4.80	.0142	3.57	.0195	4.92	.0270	6.79	.0281	7.08
13	.0173	4.36	.0073	1.83	.0138	3.47	.0174	4.39	.0202	5.09
14	.0133	3.34	.0036	0.89	.0099	2.48	.0110	2.77	.0143	3.60
15	.0058	1.46	.0017	0.42	.0071	1.80	.0068	1.70	.0099	2.48
16	.0080	2.01	.0007	0.19	.0052	1.31	.0041	1.03	.0066	1.65
17	.0019	0.47	.0003	0.08	.0038	0.96	.0024	0.61	.0041	1.04
18	.0023	0.58	.0001	0.03	.0028	0.71	.0014	0.35	.0024	0.61
19	.0028	0.70	.00005	0.01	.0021	0.53	.0008	0.20	.0012	0.31
20	.00095	0.24	.00002	0.005	.0016	0.40	.0004	0.11	.0005	0.13
21	.00014	0.34	.000007	0.002	.0012	0.30	.0002	0.06	.00015	0.04
22	.00013	0.32	.000003	0.0007	.0009	0.23	.0001	0.03	.00001	0.003
		251.78		255.21		252.57		254.72		257.02

RF = relative frequency
AF = absolute frequency

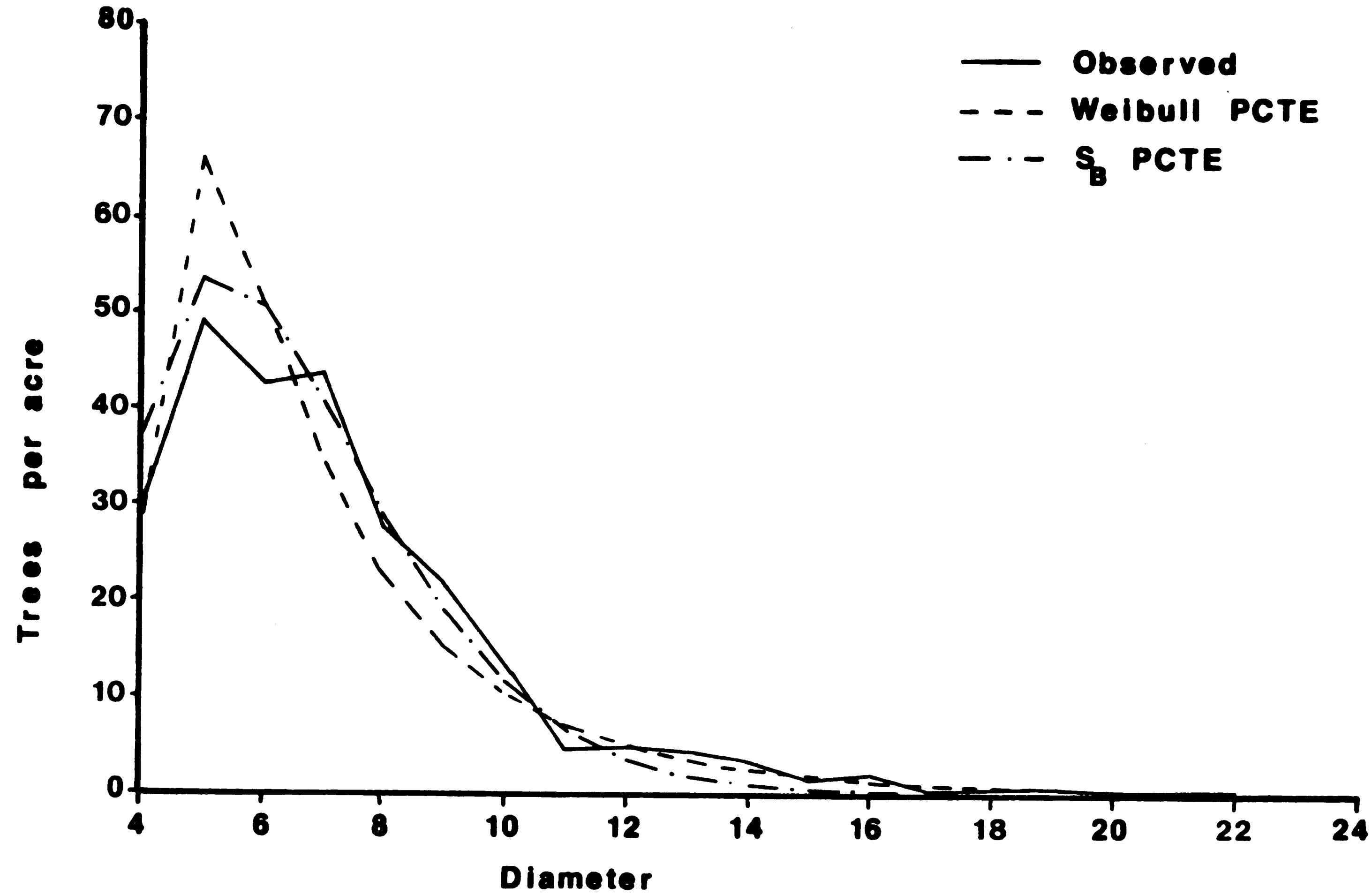

Figure 2: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 1.

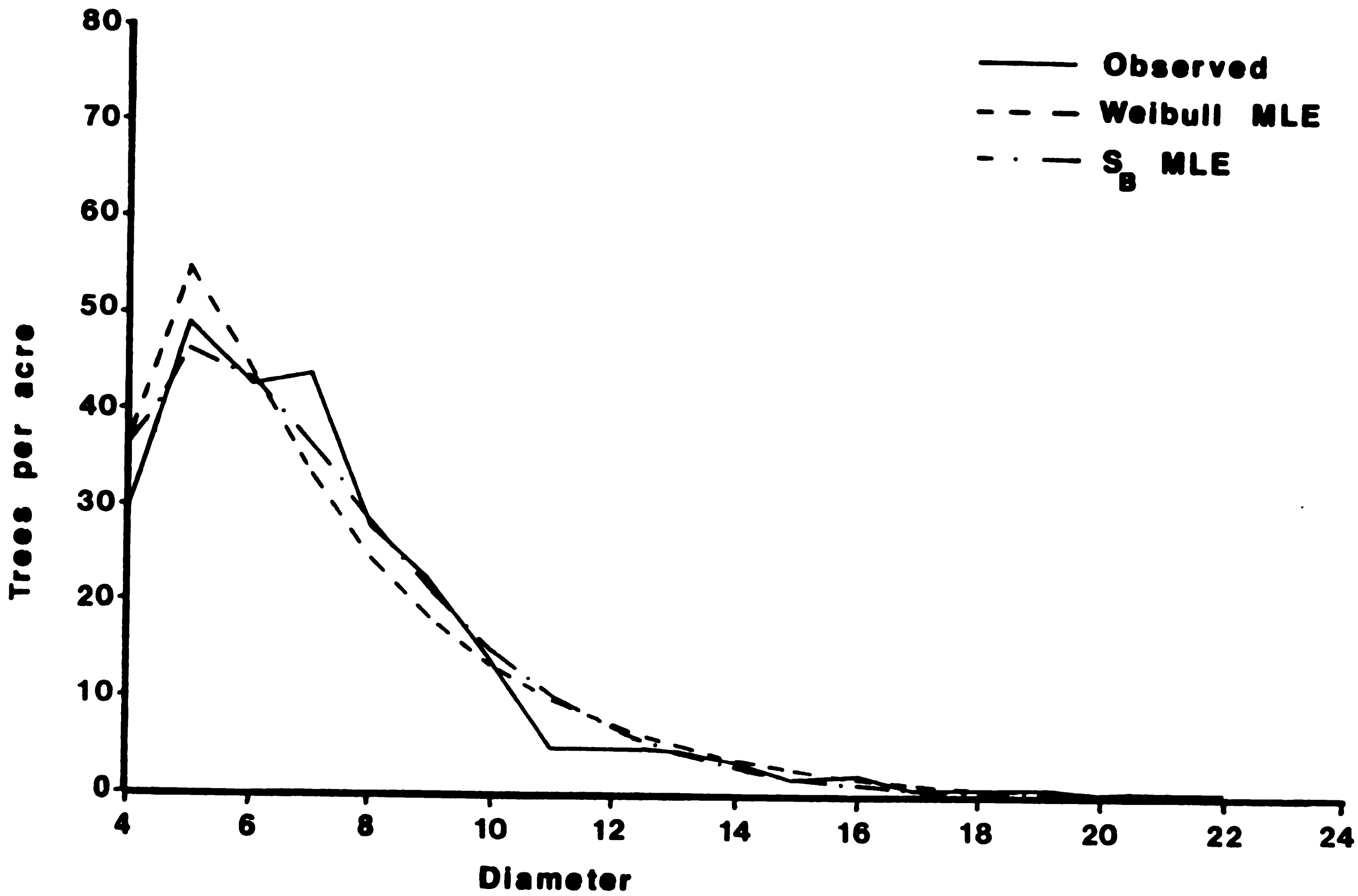

Figure 3: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 1.

Table 12 : ELTP 10: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

DBH	OBSERVED RF	OBSERVED AF	WEIBULL PCTILE RF	WEIBULL PCTILE AF	Sb PCTILE RF	Sb PCTILE AF	WEIBULL MLE RF	WEIBULL MLE AF	Sb MLE RF	Sb MLE AF
4	.1513	39.46	.1016	26.51	.1555	40.56	.1468	38.30	.1672	43.63
5	.1780	46.95	.1574	41.07	.1899	49.54	.1791	46.73	.2088	54.47
6	.1361	35.52	.1661	43.32	.1631	42.55	.1668	43.52	.1654	43.16
7	.1458	38.04	.1512	39.44	.1282	33.46	.1400	36.53	.1250	32.62
8	.0951	24.80	.1256	32.76	.0977	25.48	.1103	28.78	.0941	24.55
9	.1094	28.58	.0975	25.42	.0733	19.12	.0830	21.66	.0710	18.53
10	.0593	15.46	.0716	18.67	.0545	14.22	.0602	15.72	.0537	14.01
11	.0400	10.44	.0502	13.08	.0402	10.49	.0424	11.06	.0406	10.59
12	.0290	7.56	.0337	8.79	.0294	7.66	.0291	7.58	.0305	7.95
13	.0225	5.89	.0218	5.69	.0212	5.53	.0195	5.08	.0227	5.91
14	.0122	3.19	.0136	3.56	.0151	3.94	.0128	3.33	.0166	4.32
15	.0066	1.73	.0082	2.15	.0105	2.75	.0082	2.14	.0118	3.08
16	.0068	1.78	.0048	1.26	.0071	1.87	.0052	1.35	.0081	2.11
17	.0025	0.66	.0028	0.72	.0047	1.22	.0032	0.84	.0052	1.36
18	.0016	0.43	.0015	0.40	.0029	0.77	.0020	0.51	.0031	0.80
19	.0000	0.00	.0008	0.22	.0017	0.45	.0012	0.31	.0015	0.40
20	.0005	0.14	.0004	0.11	.0009	0.24	.0007	0.18	.0005	0.14
21	.0011	0.29	.0002	0.06	.0004	0.11	.0004	0.11	.00008	0.02
		260.89		263.25		259.97		263.73		267.66

RF = relative frequency
AF = absolute frequency

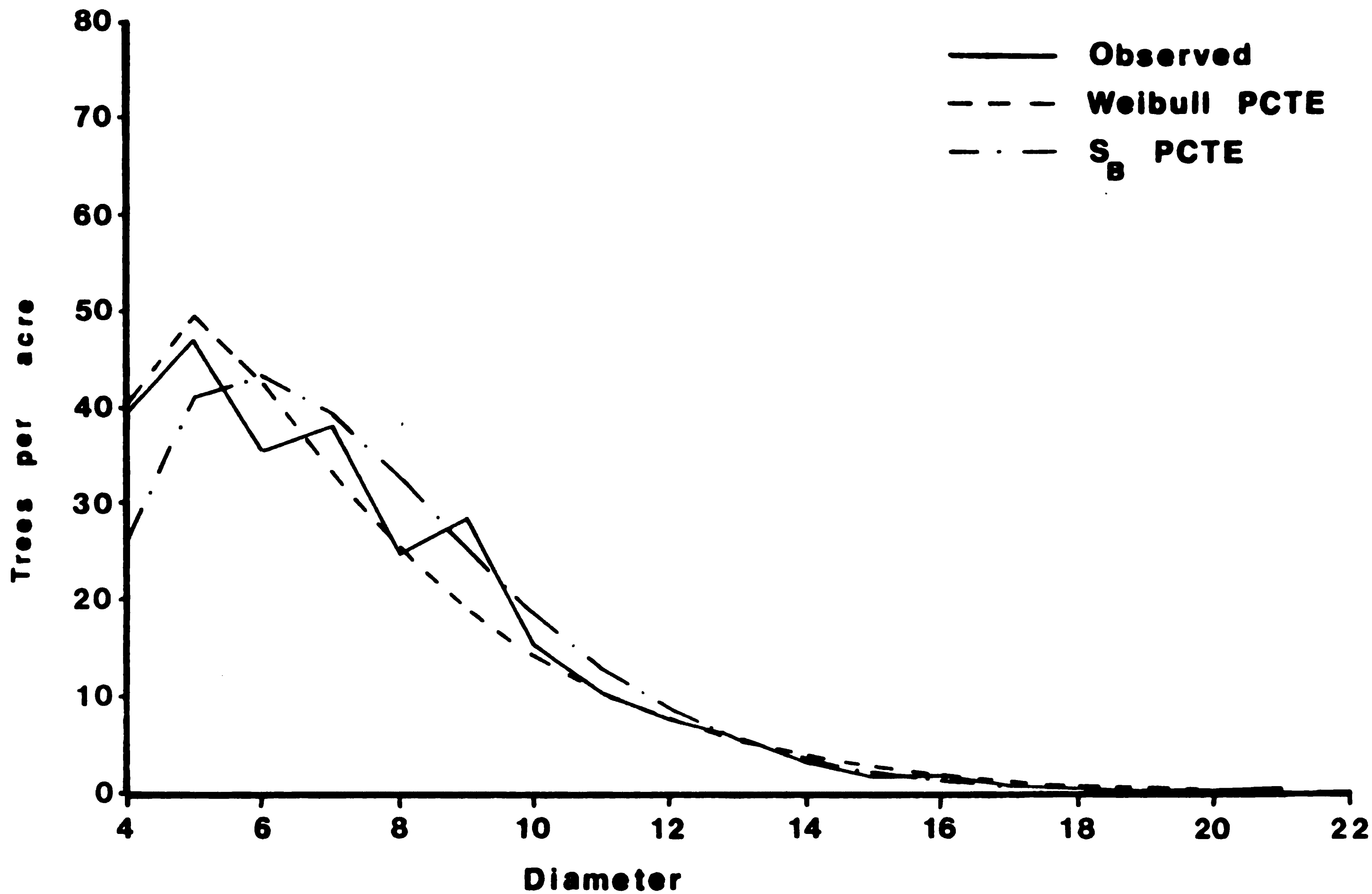

Figure 4: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 10.

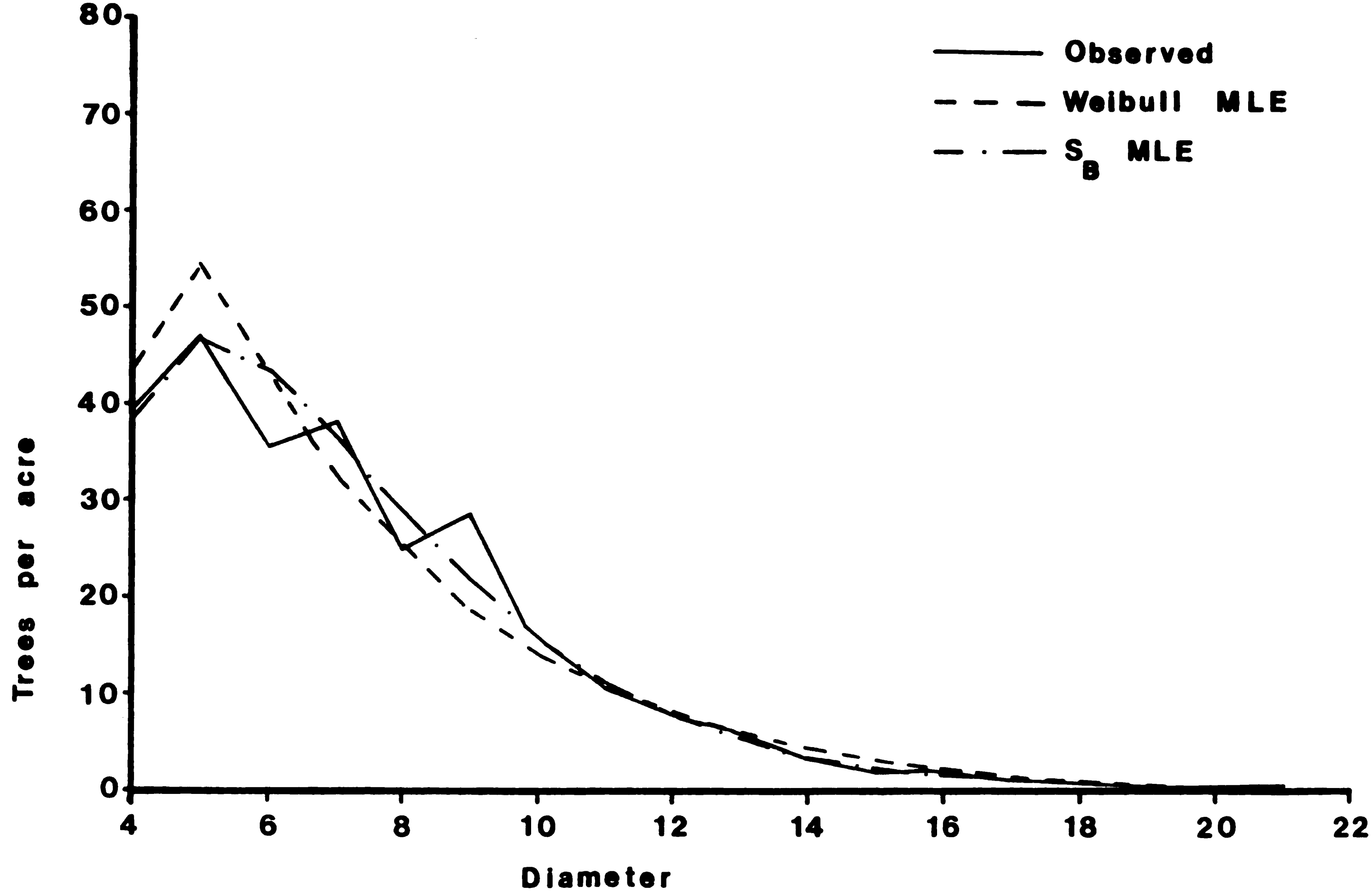

Figure 5: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 10.

Table 13 : ELTP 12: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1791	46.72	.2076	54.15	.1729	45.10	.1706	44.50	.2000	52.18
5	.1925	50.21	.1570	40.96	.2098	53.16	.1775	46.30	.2037	53.13
6	.1160	30.25	.1230	32.09	.1547	40.36	.1551	40.46	.1539	40.14
7	.1386	36.15	.0976	25.45	.1166	30.41	.1270	33.13	.1149	29.98
8	.0728	18.99	.0779	20.33	.0893	23.28	.1000	26.09	.0868	22.65
9	.0674	17.57	.0625	16.31	.0694	18.10	.0766	19.99	.0664	17.32
10	.0757	19.76	.0503	13.13	.0545	14.22	.0575	14.99	.0512	13.36
11	.0584	15.23	.0406	10.60	.0431	11.24	.0424	11.05	.0398	10.37
12	.0350	9.13	.0329	8.57	.0342	8.91	.0308	8.03	.0309	8.05
13	.0187	4.87	.0266	6.95	.0270	7.05	.0221	5.76	.0239	6.23
14	.0101	2.63	.0216	5.64	.0212	5.53	.0156	4.08	.0184	4.79
15	.0146	3.82	.0176	4.58	.0164	4.28	.0110	2.86	.0139	3.63
16	.0034	0.89	.0143	3.73	.0124	3.34	.0076	1.99	.0103	2.69
17	.0084	2.18	.0116	3.04	.0091	2.36	.0053	1.37	.0074	1.93
18	.0061	1.59	.0095	2.48	.0062	1.63	.0036	0.94	.0050	1.32
19	.0018	0.47	.0077	2.02	.0039	1.01	.0024	0.64	.0032	0.83
20	.0011	0.29	.0063	1.65	.0020	0.51	.0016	0.43	.0017	0.45
21	.0000	0.00	.0052	1.35	.0006	0.15	.0011	0.29	.0007	0.19
22	.0005	0.12	.0042	1.10	.000006	0.0015	.0007	0.19	.0001	0.03
		260.87		254.12		270.55		263.08		269.25

RF = relative frequency
AF = absolute frequency

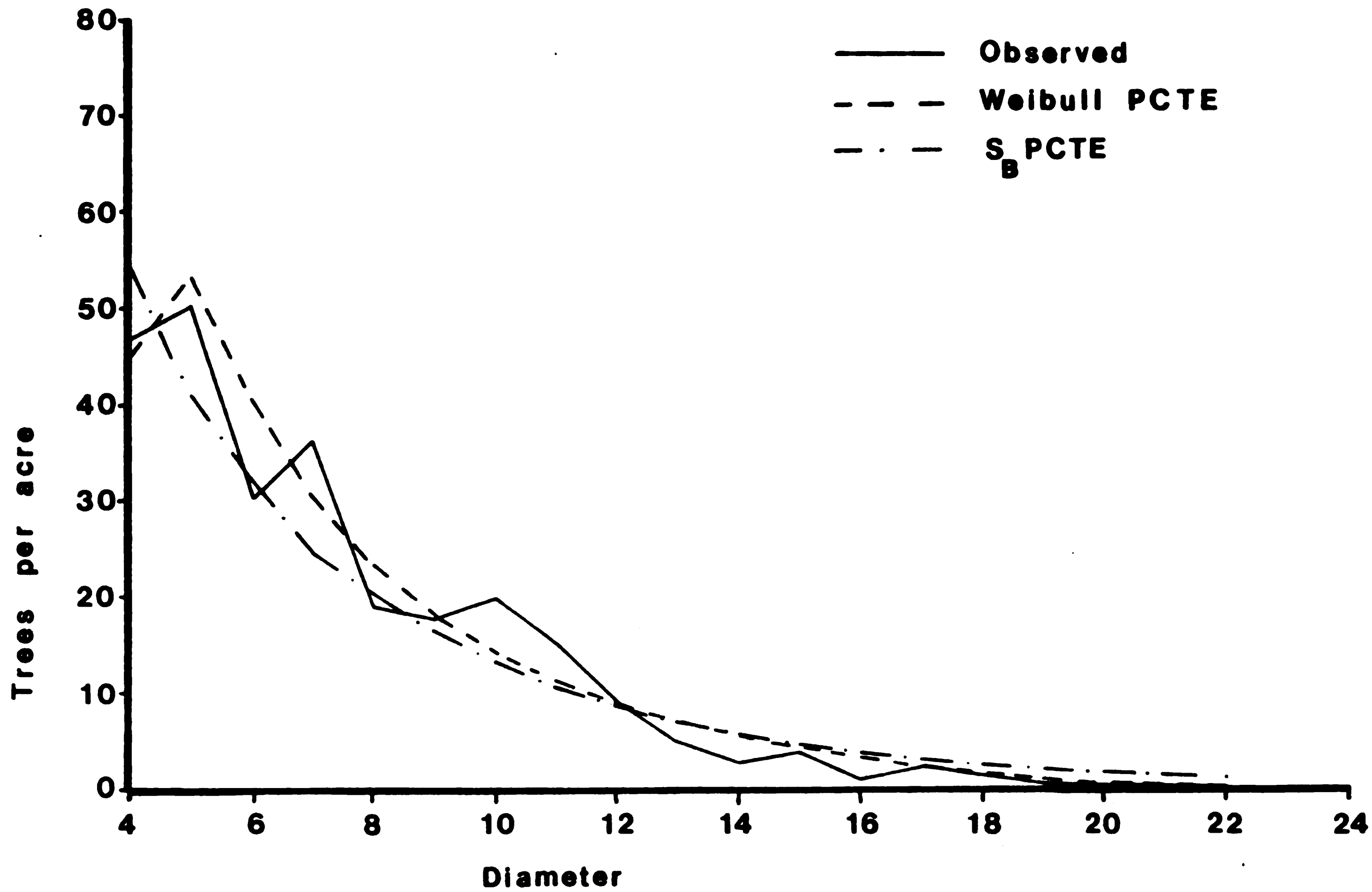

Figure 6: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 12.

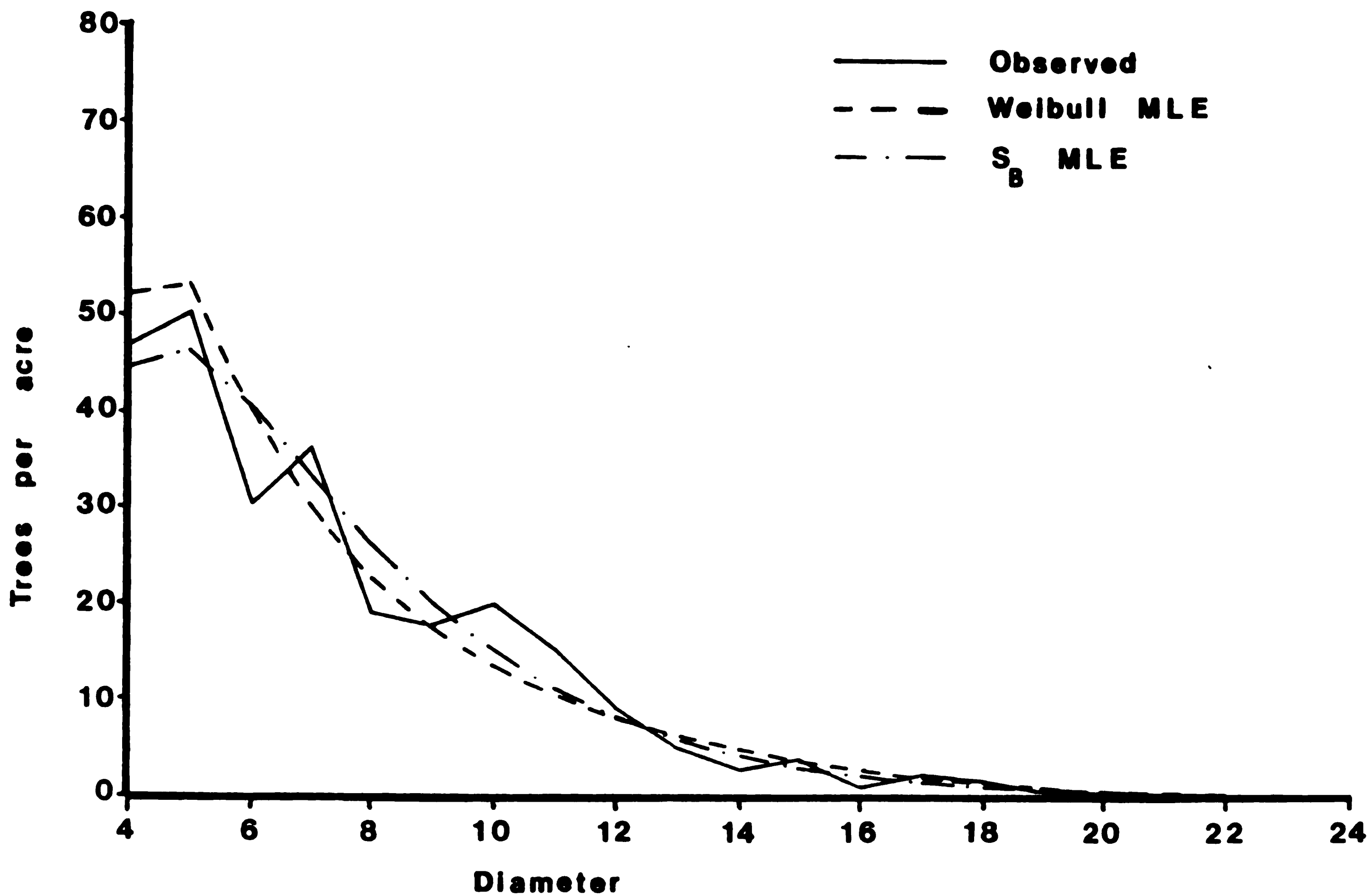

Figure 7: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 12.

Table 14 : ELTP 20: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1019	23.87	.0928	21.73	.0930	21.96	.0964	22.58	.1173	27.47
5	.1435	33.61	.1282	30.04	.1350	31.80	.1347	31.56	.1614	37.80
6	.1359	31.83	.1335	31.28	.1294	30.32	.1399	32.78	.1414	33.12
7	.0732	17.15	.1255	29.40	.1140	26.69	.1304	30.54	.1171	27.44
8	.1478	34.62	.1111	26.03	.0979	22.93	.1140	26.69	.0961	22.51
9	.0886	20.75	.0943	22.08	.0833	19.52	.0951	22.29	.0788	18.47
10	.0652	15.28	.0774	18.13	.0706	16.53	.0766	17.95	.0648	15.18
11	.0593	13.89	.0619	14.49	.0595	13.94	.0599	14.04	.0533	12.48
12	.0498	11.67	.0483	11.31	.0500	11.70	.0457	10.70	.0437	10.24
13	.0656	15.37	.0369	8.65	.0416	9.75	.0340	7.97	.0357	8.37
14	.0250	5.85	.0277	6.50	.0344	8.05	.0248	5.82	.0290	6.78
15	.0145	3.40	.0205	4.80	.0280	6.55	.0178	4.17	.0232	5.43
16	.0064	1.49	.0149	3.49	.0223	5.22	.0125	2.94	.0182	4.26
17	.0091	2.11	.0107	2.50	.0173	4.04	.0087	2.04	.0138	3.24
18	.0108	2.36	.0076	1.77	.0128	3.00	.0059	1.39	.0101	2.36
19	.0027	0.63	.0053	1.24	.0089	2.08	.0040	0.93	.0068	1.60
20	.0008	0.19	.0036	0.85	.0055	1.28	.0026	0.62	.0041	0.96
21	.0000	0.00	.0025	0.58	.0027	0.63	.0017	0.41	.0019	0.45
22	.0006	0.16	.0017	0.39	.0007	0.17	.0011	0.26	.0004	0.10
		234.23		235.26		236.17		235.67		238.26

RF = relative frequency
AF = absolute frequency

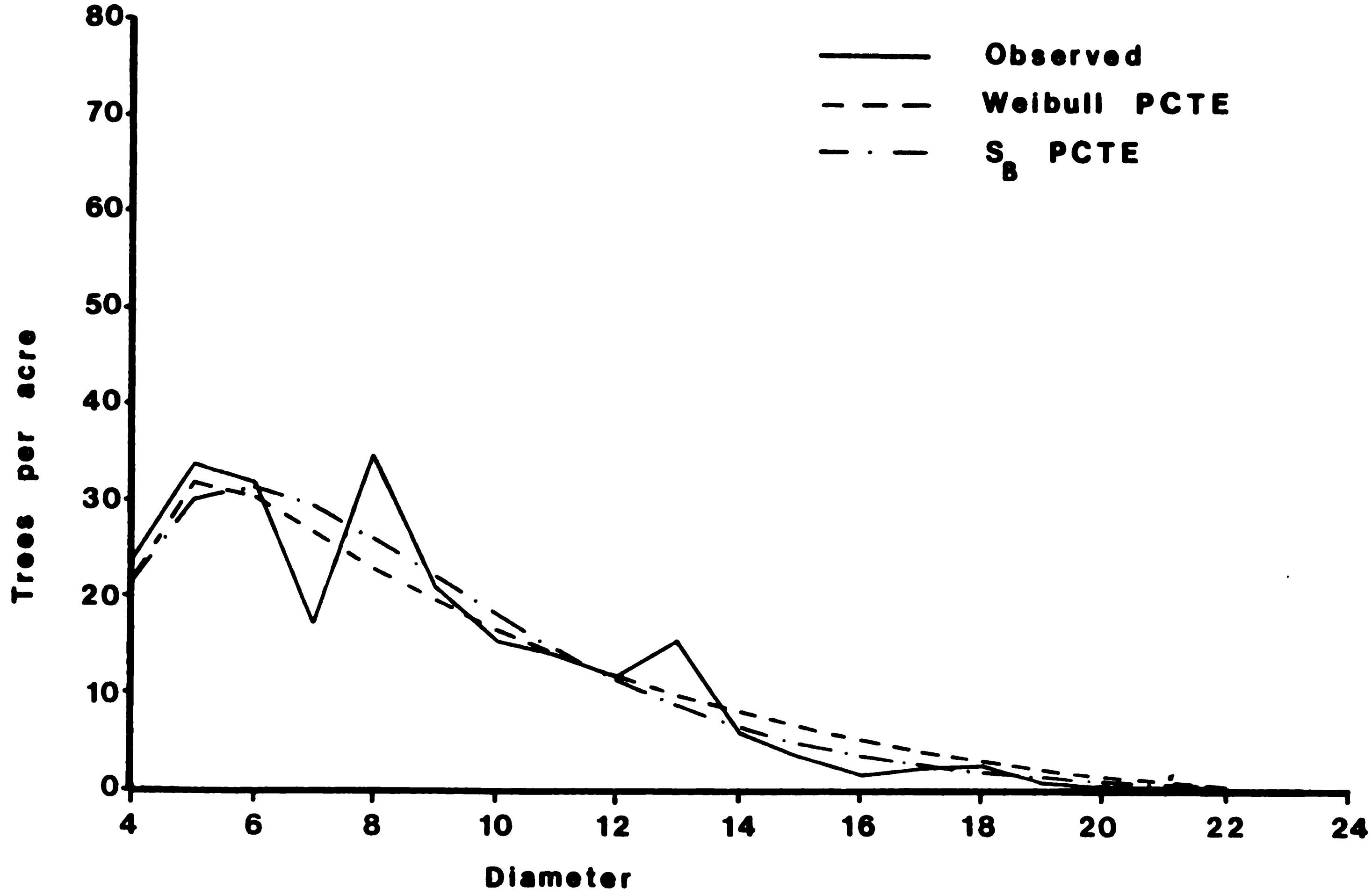

Figure 8: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 20.

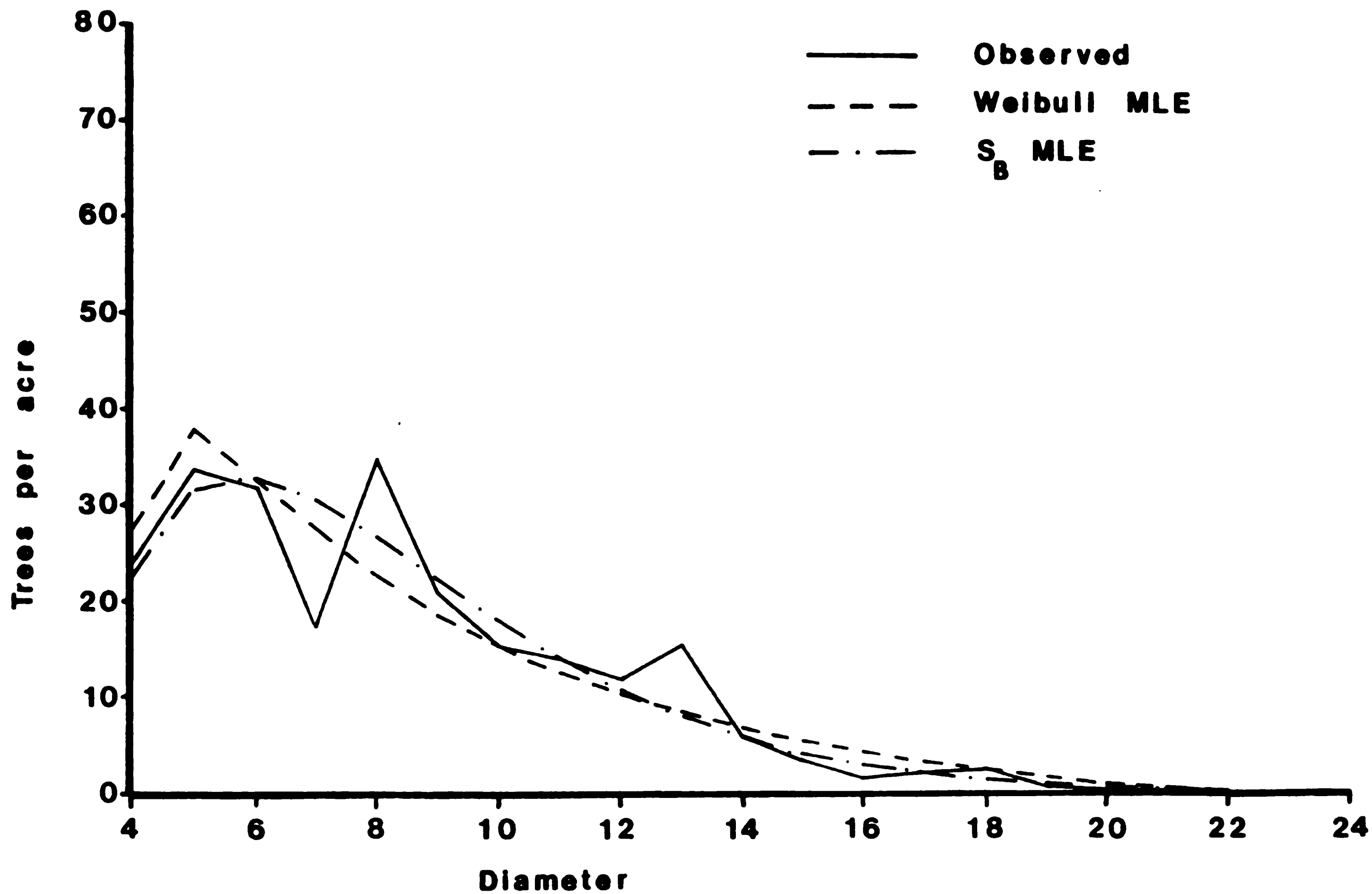

Figure 9: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 20.

Table 15 : ELTP 21: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1463	44.56	.1122	34.18	.1808	55.05	.1587	48.32	.1721	52.41
5	.1940	59.08	.1598	48.68	.2374	42.29	.1752	53.37	.2088	63.60
6	.1394	42.44	.1624	49.47	.1726	52.56	.1574	47.95	.1623	49.42
7	.1433	43.65	.1452	44.22	.1204	36.67	.1309	39.86	.1210	36.85
8	.1176	35.81	.1200	36.53	.0848	25.82	.1038	31.62	.0903	27.51
9	.0619	18.86	.0936	28.49	.0606	18.47	.0797	24.27	.0680	20.70
10	.0569	17.32	.0697	21.22	.0441	13.42	.0596	18.16	.0515	15.68
11	.0304	9.26	.0499	15.19	.0324	9.87	.0437	13.30	.0392	11.95
12	.0290	8.84	.0345	10.51	.0241	7.34	.0314	9.56	.0300	9.12
13	.0188	5.73	.0232	7.06	.0181	5.51	.0222	6.77	.0229	6.96
14	.0222	6.76	.0151	4.61	.0136	4.15	.0155	4.73	.0174	5.29
15	.0156	4.75	.0096	2.93	.0103	3.15	.0107	3.26	.0131	3.99
16	.0072	2.19	.0060	1.82	.0078	2.39	.0073	2.22	.0098	2.97
17	.0064	1.94	.0036	1.10	.0060	1.81	.0049	1.49	.0072	2.18
18	.0036	1.10	.0022	0.66	.0045	1.38	.0033	1.00	.0051	1.56
19	.0033	0.99	.0013	0.38	.0034	1.04	.0022	0.66	.0035	1.08
20	.0029	0.89	.0007	0.22	.0026	0.78	.0014	0.43	.0023	0.71
21	.0000	0.00	.0004	0.12	.0019	0.59	.0009	0.28	.0014	0.43
22	.00036	0.11	.0002	0.07	.0014	0.44	.0006	0.18	.0008	0.24
23	.00032	0.10	.0001	0.04	.0010	0.32	.0004	0.12	.0004	0.11
24	.0003	0.09	.00006	0.02	.0008	0.23	.0002	0.07	.0001	0.03
25	.00026	0.08	.00003	0.01	.0005	0.16	.00015	0.05	.00001	0.004
		304.55		307.52		313.45		307.66		312.79

RF = relative frequency
AF = absolute frequency

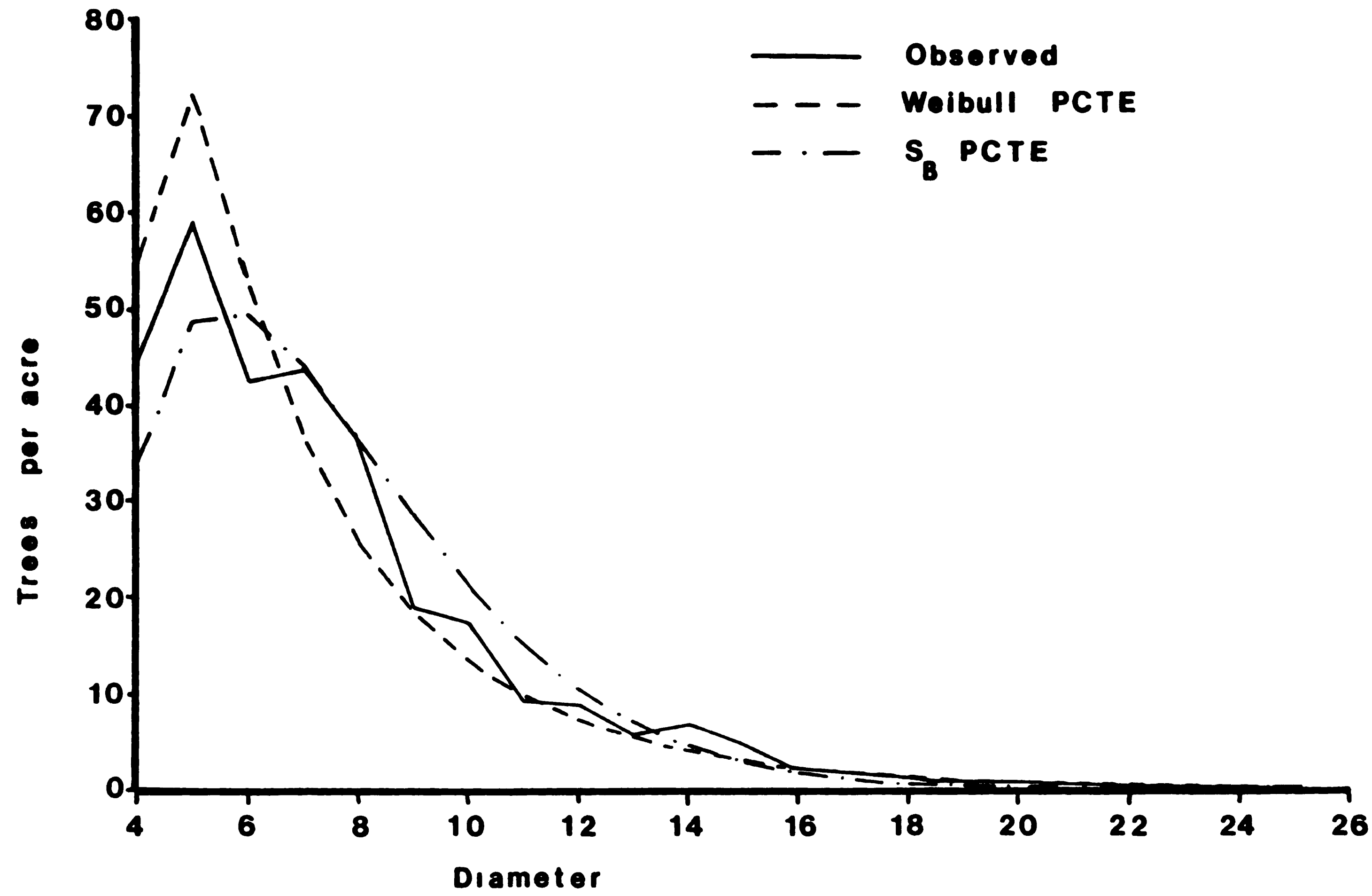

Figure 10: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 21.

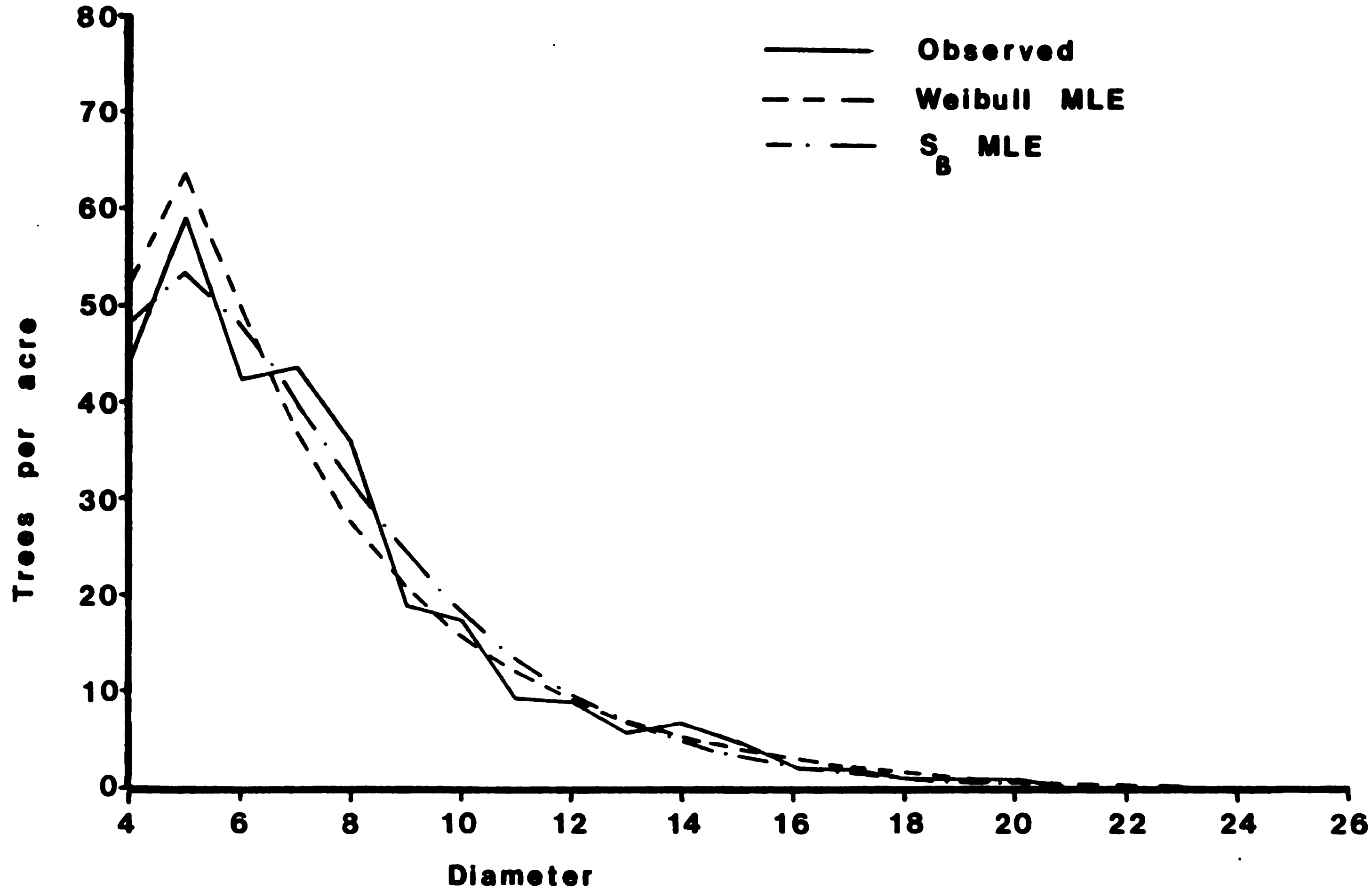

Figure 11: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 21.

Table 16 : ELTP 35: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1023	26.99	.0928	24.48	.1022	26.97	.1118	29.49	.1200	34.65
5	.2125	56.08	.1282	33.84	.1584	41.80	.1393	36.75	.1703	44.92
6	.0718	18.95	.1336	35.29	.1359	35.86	.1372	36.21	.1465	38.64
7	.0873	23.04	.1255	33.12	.1110	29.30	.1242	32.77	.1184	31.23
8	.1251	33.00	.1111	29.32	.0906	23.91	.1069	28.22	.0947	24.99
9	.0551	14.54	.0943	24.88	.0744	19.64	.0889	23.47	.0759	20.03
10	.0886	23.37	.0774	20.43	.0616	16.26	.0720	19.00	.0611	16.11
11	.0695	18.35	.0619	16.32	.0513	13.54	.0570	15.04	.0493	13.00
12	.0564	14.88	.0483	12.75	.0429	11.33	.0443	11.69	.0398	10.51
13	.0449	11.84	.0369	9.75	.0360	9.51	.0339	8.94	.0322	8.49
14	.0279	7.35	.0277	7.32	.0303	7.99	.0255	6.73	.0260	6.86
15	.0179	4.71	.0206	5.41	.0254	6.70	.0190	5.01	.0208	5.50
16	.0173	4.56	.0149	3.99	.0213	5.61	.0139	3.68	.0166	4.38
17	.0061	1.61	.0107	2.82	.0177	4.67	.0101	2.67	.0131	3.45
18	.0063	1.65	.0076	1.99	.0146	3.86	.0073	1.92	.0102	2.68
19	.0041	1.07	.0053	1.39	.0119	3.15	.0052	1.37	.0077	2.03
20	.0022	0.59	.0036	0.96	.0096	2.53	.0037	0.96	.0057	1.50
21	.0017	0.46	.0025	0.65	.0075	1.98	.0026	0.67	.0040	1.06
22	.0012	0.32	.0017	0.44	.0057	1.50	.0018	0.47	.0027	0.70
23	.00076	0.20	.0011	0.29	.0041	1.08	.0012	0.32	.0016	0.42
24	.0004	0.11	.0007	0.19	.0027	0.71	.0008	0.22	.0008	0.22
25	.0004	0.11	.0005	0.13	.0015	0.40	.0006	0.15	.0003	0.08
26	.0003	0.08	.0003	0.08	.0006	0.17	.0004	0.10	.00004	0.01
		263.86		265.72		268.47		265.84		268.47

RF = relative frequency
AF = absolute frequency

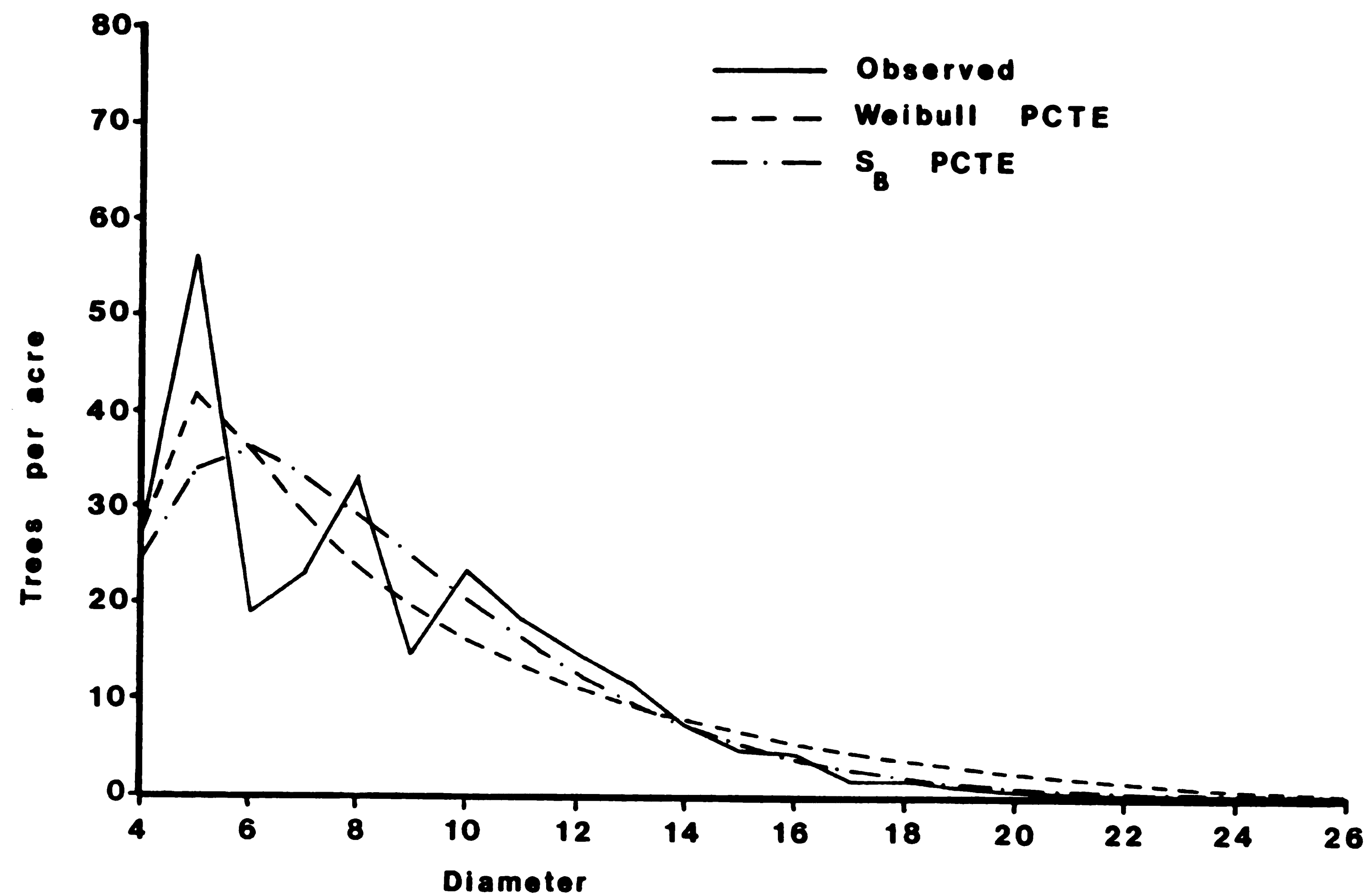

Figure 12: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 35.

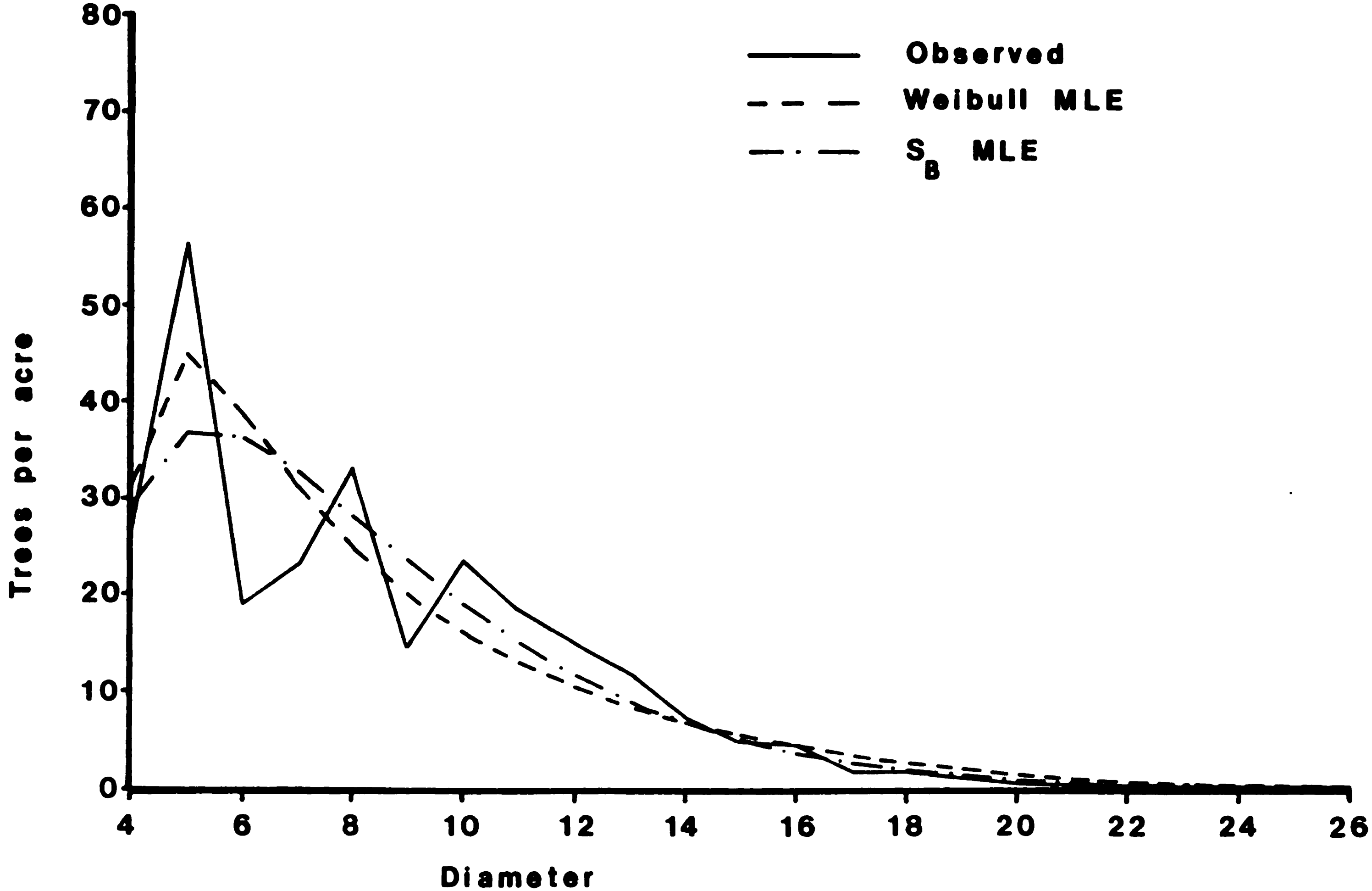

Figure 13: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 35.

Table 17 : ELTP 37: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1571	28.19	.0723	12.98	.1054	18.92	.0689	12.37	.1013	18.19
5	.0613	11.00	.0935	16.78	.1142	20.51	.0971	17.42	.1289	23.14
6	.0242	4.34	.0980	17.58	.1044	18.73	.1052	18.88	.1145	20.55
7	.0496	8.88	.0956	17.16	.0920	16.50	.1041	18.68	.0977	17.54
8	.0559	10.03	.0897	16.10	.0803	14.41	.0979	17.57	.0832	14.93
9	.0764	13.72	.0819	14.70	.0701	12.58	.0889	15.96	.0712	12.77
10	.1440	25.85	.0733	13.16	.0612	10.99	.0786	14.11	.0612	10.99
11	.0760	13.65	.0646	11.60	.0536	9.62	.0681	12.22	.0530	9.51
12	.0973	17.46	.0562	10.09	.0470	8.43	.0579	10.39	.0461	8.27
13	.0705	12.65	.0483	8.68	.0412	7.40	.0485	8.70	.0402	7.21
14	.0587	10.54	.0412	7.39	.0361	6.49	.0400	7.18	.0351	3.60
15	.0436	7.83	.0347	6.24	.0316	5.68	.0326	5.86	.0307	5.51
16	.0262	4.71	.0291	5.22	.0276	4.96	.0263	4.72	.0268	4.81
17	.0310	5.57	.0242	4.34	.0240	4.31	.0210	3.77	.0234	4.20
18	.0092	1.65	.0200	3.59	.0208	3.73	.0166	2.97	.0203	3.65
19	.0051	0.91	.0164	2.94	.0178	3.19	.0130	2.33	.0175	3.15
20	.0059	1.06	.0134	2.40	.0151	2.70	.0101	1.81	.0150	2.70
21	.0023	0.41	.0108	1.95	.0125	2.25	.0077	1.39	.0127	2.28
22	.0011	0.20	.0088	1.57	.0102	1.83	.0059	1.06	.0105	1.89
23	.0011	0.20	.0070	1.26	.0080	1.44	.0045	0.80	.0085	1.53
24	.0009	0.16	.0056	1.01	.0060	1.07	.0034	0.61	.0066	1.19
25	.0000	0.00	.0045	0.80	.0041	0.73	.0025	0.45	.0049	0.87
26	.0000	0.00	.0035	0.64	.0024	0.43	.0019	0.34	.0032	0.58
27	.0014	0.25	.0028	0.50	.0010	0.17	.0014	0.25	.0017	0.31
28	.0013	0.23	.0022	0.39	.0001	0.02	.0010	0.18	.0005	0.09
		179.49		179.06		177.09		180.02		182.16

RF = relative frequency
AF = absolute frequency

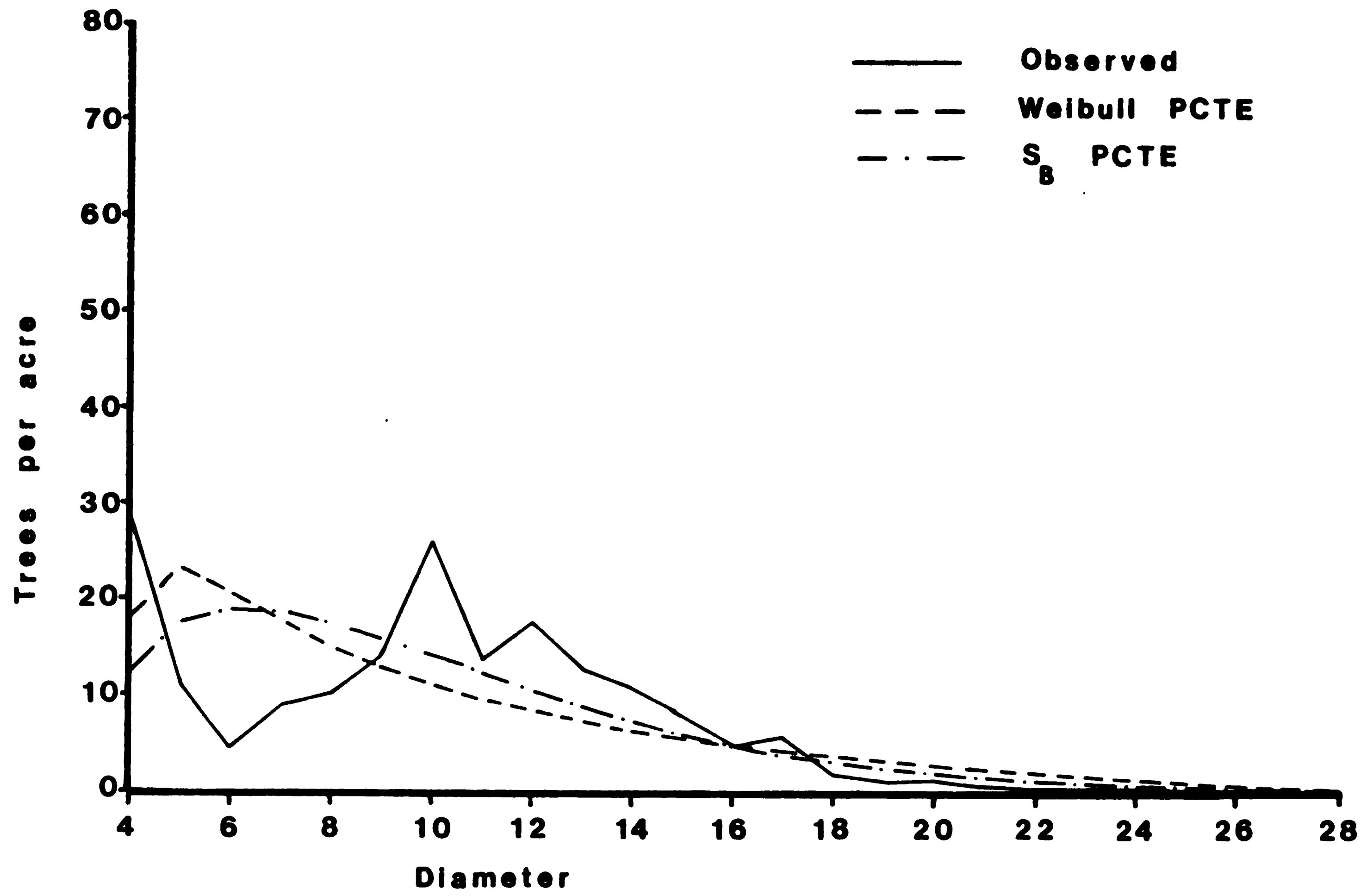

Figure 14: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 37.

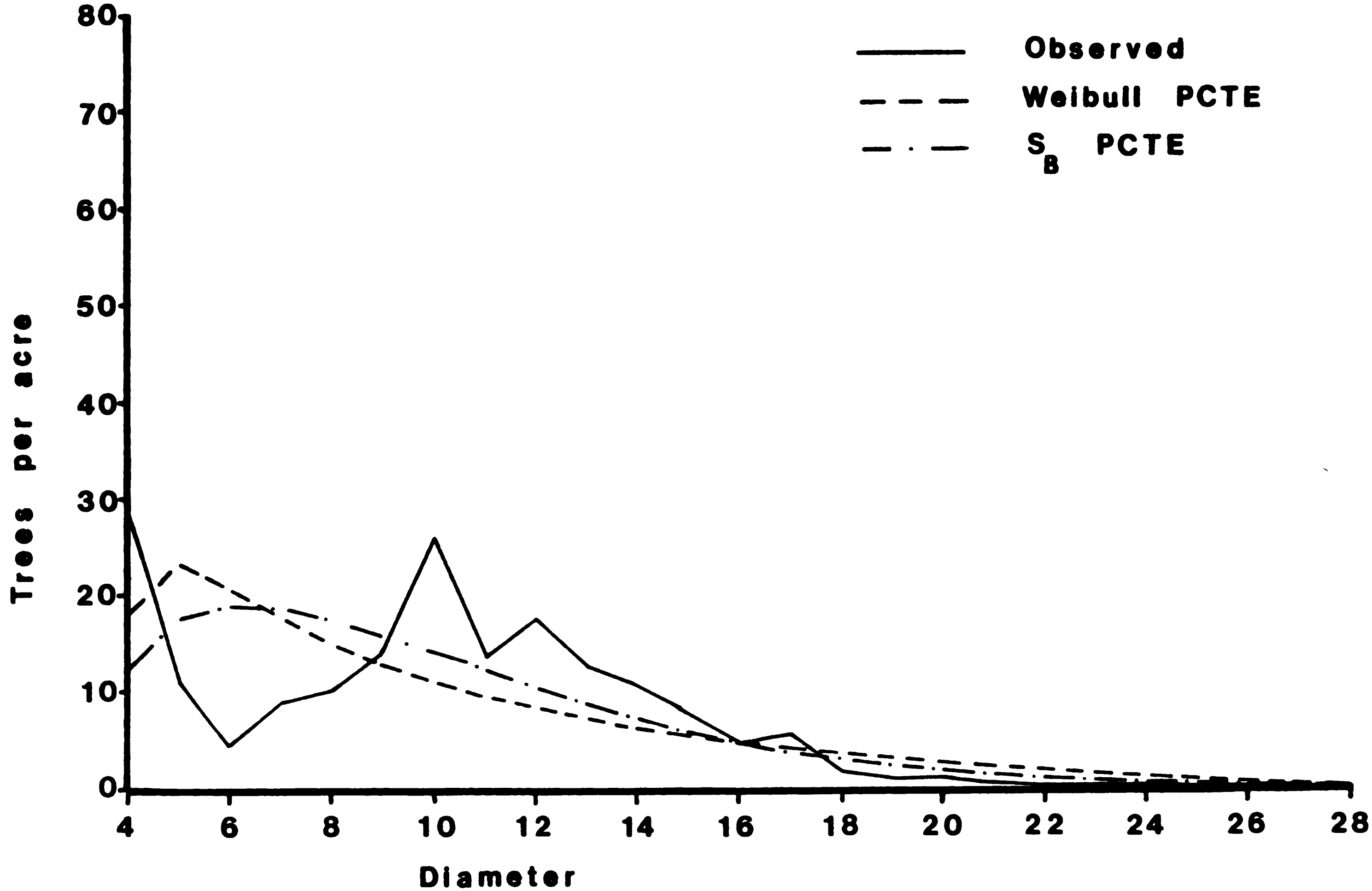

Figure 15: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 37.

Table 18 : ELTP 40: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

	OBSERVED		WEIBULL PCTILE		Sb PCTILE		WEIBULL MLE		Sb MLE	
DBH	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF
4	.1794	55.77	.2023	62.90	.2036	63.30	.1575	48.99	.1883	58.54
5	.1577	49.05	.1584	49.26	.2030	63.13	.1787	55.58	.2032	63.18
6	.1508	46.90	.1258	39.11	.1527	47.47	.1616	50.26	.1572	48.89
7	.1039	32.30	.1003	31.20	.1145	35.61	.1341	41.70	.1190	37.01
8	.0997	31.01	.0803	24.96	.0872	27.13	.1056	32.84	.0906	28.18
9	.0961	29.87	.0643	20.00	.0674	20.95	.0802	24.92	.0696	21.63
10	.0782	24.32	.0516	16.05	.0525	16.32	.0591	18.37	.0536	16.78
11	.0390	12.14	.0415	12.89	.0411	12.77	.0425	13.21	.0414	12.87
12	.0361	11.24	.0333	10.36	.0321	9.99	.0299	9.31	.0318	9.89
13	.0172	5.36	.0268	8.34	.0250	7.76	.0207	6.44	.0242	7.52
14	.0246	7.64	.0216	6.71	.0191	5.95	.0141	4.39	.0181	5.61
15	.0075	2.39	.0174	5.41	.0144	4.47	.0095	2.94	.0131	4.07
16	.0058	1.79	.0140	4.36	.0104	3.23	.0063	1.95	.0091	2.82
17	.0022	0.68	.0113	3.51	.0071	2.21	.0041	1.27	.0058	1.81
18	.0007	0.21	.0091	2.83	.0044	1.36	.0026	0.82	.0033	1.02
19	.0006	0.18	.0074	2.29	.0022	0.68	.0017	0.52	.0014	0.44
20	.0005	0.16	.0060	1.85	.0006	0.20	.0011	0.33	.0003	0.09
		310.95		302.02		322.51		313.84		320.27

RF = relative frequency
AF = absolute frequency

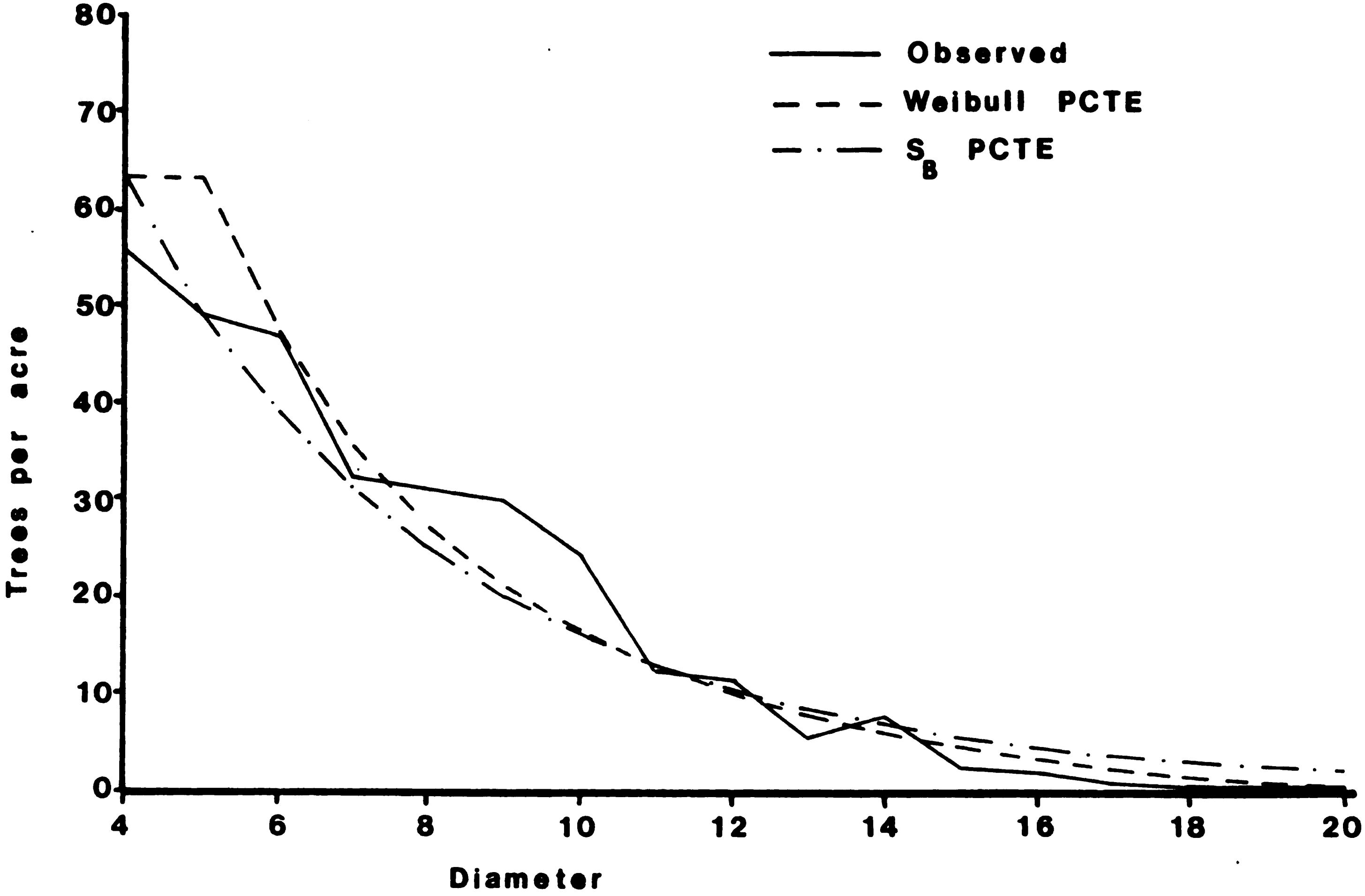

Figure 16: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 40.

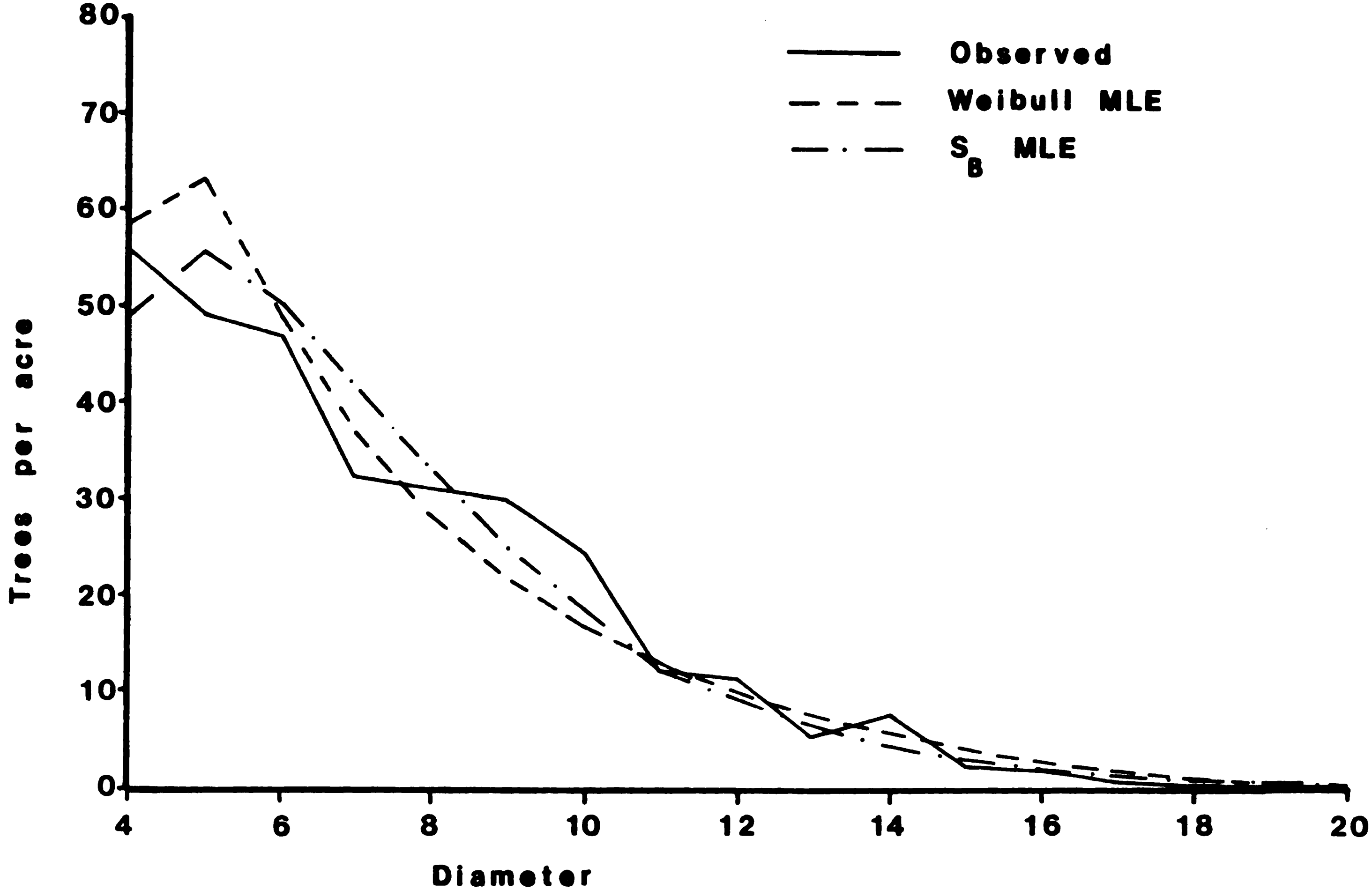

Figure 17: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 40.

Table 19 : ELTP 43: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

DBH	OBSERVED RF	OBSERVED AF	WEIBULL PCTILE RF	WEIBULL PCTILE AF	Sb PCTILE RF	Sb PCTILE AF	WEIBULL MLE RF	WEIBULL MLE AF	Sb MLE RF	Sb MLE AF
4	.2185	57.30	.2190	57.45	.2096	54.97	.1710	44.85	.2091	54.84
5	.1165	30.56	.1531	40.15	.1706	44.76	.1578	41.40	.1816	47.62
6	.1213	31.83	.1164	30.53	.1257	32.97	.1344	35.26	.1345	35.29
7	.0719	18.71	.0912	23.91	.0956	25.08	.1112	29.18	.1016	26.64
8	.1229	32.23	.0726	19.03	.0750	19.66	.0906	23.75	.0786	20.62
9	.0647	16.98	.0584	15.31	.0602	15.79	.0729	19.12	.0621	16.29
10	.0670	17.57	.0473	12.41	.0492	12.90	.0582	15.26	.0498	13.08
11	.0482	12.63	.0386	10.13	.0408	10.69	.0461	12.10	.0405	10.63
12	.0364	9.55	.0316	8.30	.0341	8.95	.0364	9.54	.0332	8.71
13	.0310	8.14	.0260	6.83	.0288	7.56	.0285	7.48	.0274	7.19
14	.0327	8.57	.0215	5.64	.0245	6.42	.0223	5.85	.0227	5.95
15	.0091	2.38	.0178	4.67	.0209	5.47	.0174	4.55	.0188	4.94
16	.0091	2.39	.0148	3.88	.0178	4.68	.0135	3.53	.0156	4.10
17	.0161	4.23	.0129	9.29	.0152	4.00	.0104	2.73	.0129	3.39
18	.0153	4.01	.0103	2.69	.0130	3.41	.0080	2.11	.0106	2.78
19	.0081	2.12	.0086	2.25	.0111	2.90	.0062	1.62	.0086	2.27
20	.0022	0.57	.0072	1.88	.0094	2.45	.0048	1.25	.0069	1.82
21	.0006	0.17	.0060	1.58	.0078	2.05	.0036	0.95	.0055	1.43
22	.0030	0.79	.0050	1.32	.0065	1.70	.0028	0.73	.0042	1.09
23	.0027	0.72	.0042	1.11	.0052	1.38	.0021	0.56	.0030	0.80
24	.0015	0.40	.0036	0.94	.0041	1.08	.0016	0.42	.0021	0.54
25	.0009	0.24	.0030	0.79	.0031	0.81	.0012	0.32	.0013	0.33
26	.0004	0.11	.0025	0.67	.0022	0.57	.0009	0.24	.0006	0.16
27	.00038	0.10	.0021	0.56	.0013	0.35	.0007	0.19	.0001	0.04
		262.30		255.24		270.62		263.01		270.52

RF = relative frequency
AF = absolute frequency

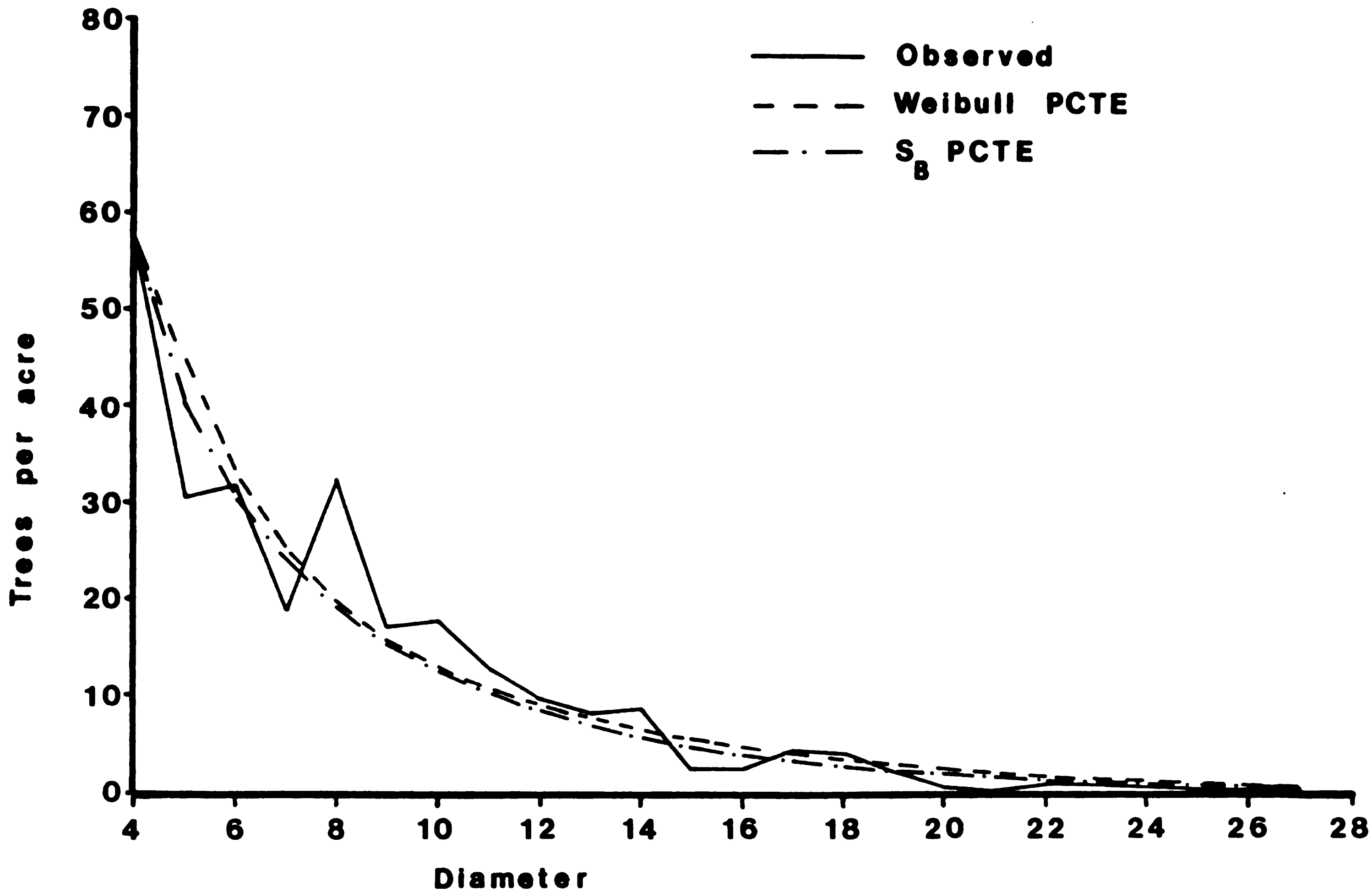

Figure 18: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 43.

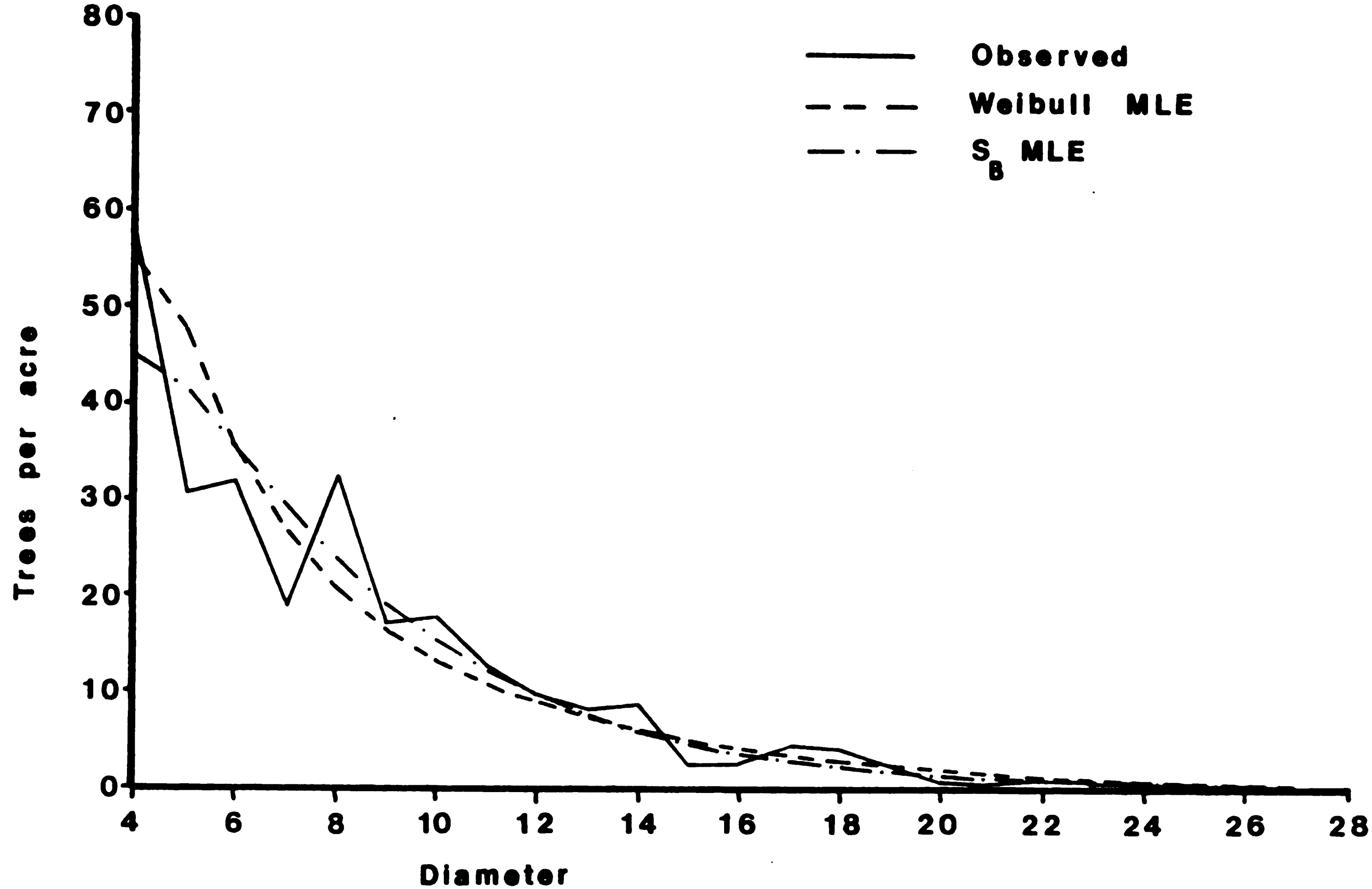

Figure 19: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 43.

Table 20 : ELTP 45: OBSERVED AND PREDICTED DIAMETER DISTRIBUTIONS

DBH	OBSERVED RF	OBSERVED AF	WEIBULL PCTILE RF	WEIBULL PCTILE AF	Sb PCTILE RF	Sb PCTILE AF	WEIBULL MLE RF	WEIBULL MLE AF	Sb MLE RF	Sb MLE AF
4	.1200	33.04	.1003	27.86	.0964	26.77	.1127	31.31	.1141	31.68
5	.1367	37.95	.1306	36.26	.1259	34.95	.1393	38.69	.1777	49.34
6	.0943	26.18	.1322	36.72	.1251	34.73	.1368	37.99	.1518	42.17
7	.1596	44.31	.1224	33.99	.1120	31.11	.1236	34.33	.1201	33.35
8	.0862	23.93	.1075	29.85	.0956	26.56	.1064	29.54	.0938	26.05
9	.0895	24.86	.0910	25.27	.0797	22.13	.0885	24.58	.0734	20.39
10	.0777	21.57	.0749	20.79	.0655	18.19	.0717	19.91	.0578	16.06
11	.0758	21.06	.0602	16.72	.0535	14.85	.0569	15.79	.0458	12.73
12	.0402	11.16	.0474	13.18	.0434	12.07	.0443	12.30	.0365	10.14
13	.0436	12.11	.0368	10.21	.0352	9.78	.0339	9.42	.0293	8.12
14	.0205	5.68	.0280	7.79	.0285	7.92	.0256	7.12	.0235	6.53
15	.0170	4.73	.0211	5.86	.0230	6.40	.0191	5.31	.0189	5.26
16	.0130	3.62	.0157	4.35	.0186	5.17	.0141	3.92	.0153	4.24
17	.0086	2.38	.0115	3.19	.0150	4.17	.0103	2.85	.0123	3.42
18	.0026	0.71	.0083	2.32	.0121	3.36	.0074	2.06	.0099	2.76
19	.00076	0.21	.0060	1.66	.0098	2.71	.0053	1.47	.0080	2.22
20	.0021	0.58	.0043	1.18	.0078	2.18	.0038	1.04	.0064	1.77
21	.0025	0.69	.0030	0.83	.0063	1.75	.0026	0.73	.0051	1.41
22	.0017	0.48	.0021	0.58	.0050	1.40	.0018	0.51	.0040	1.12
23	.0021	0.58	.0015	0.40	.0040	1.12	.0013	0.35	.0032	0.88
24	.0024	0.67	.0010	0.28	.0032	0.89	.0009	0.25	.0025	0.68
25	.0018	0.49	.0007	0.19	.0025	0.71	.0006	0.17	.0019	0.52
26	.0008	0.23	.0005	0.13	.0020	0.56	.0004	0.11	.0014	0.39
27	.0000	0.00	.0003	0.09	.0016	0.44	.0003	0.08	.0010	0.29
28	.0007	0.19	.0002	0.06	.0012	0.34	.0002	0.05	.0007	0.21
29	.0000	0.00	.0001	0.04	.0010	0.27	.0001	0.03	.0005	0.14
30	.0003	0.08	.00009	0.025	.00075	0.21	.00008	0.02	.0003	0.09
31	.0003	0.08	.00006	0.02	.0006	0.16	.00005	0.01	.0002	0.06
32	.0000	0.00	.00004	0.01	.0004	0.12	.00003	0.009	.0001	0.03
33	.0003	0.08	.000025	0.007	.0003	0.09	.00002	0.006	.00006	0.02
34	.0000	0.00	.000016	0.004	.0002	0.07	.00001	0.004	.00002	0.01
35	.0000	0.00	.00001	0.003	.00018	0.05	.000009	0.003	.000005	0.001
36	.0002	0.06	.000006	0.002	.00013	0.04	.000006	0.002	.0000003	0.000
		277.71		279.85		271.25		279.99		282.07

RF = relative frequency AF = absolute frequency

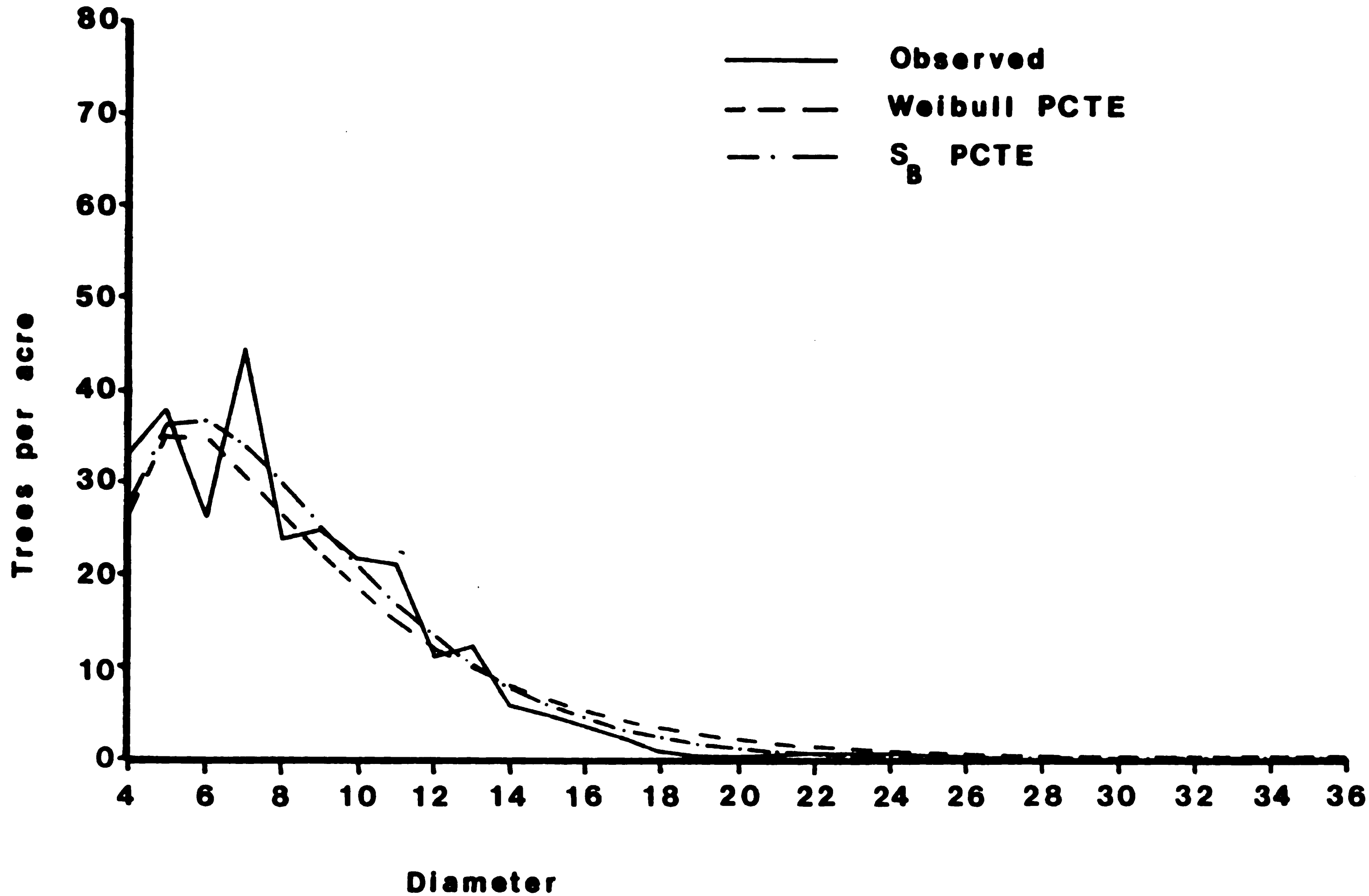

Figure 20: Observed and predicted (percentile parameter estimation) diameter distributions for ELTP 45.

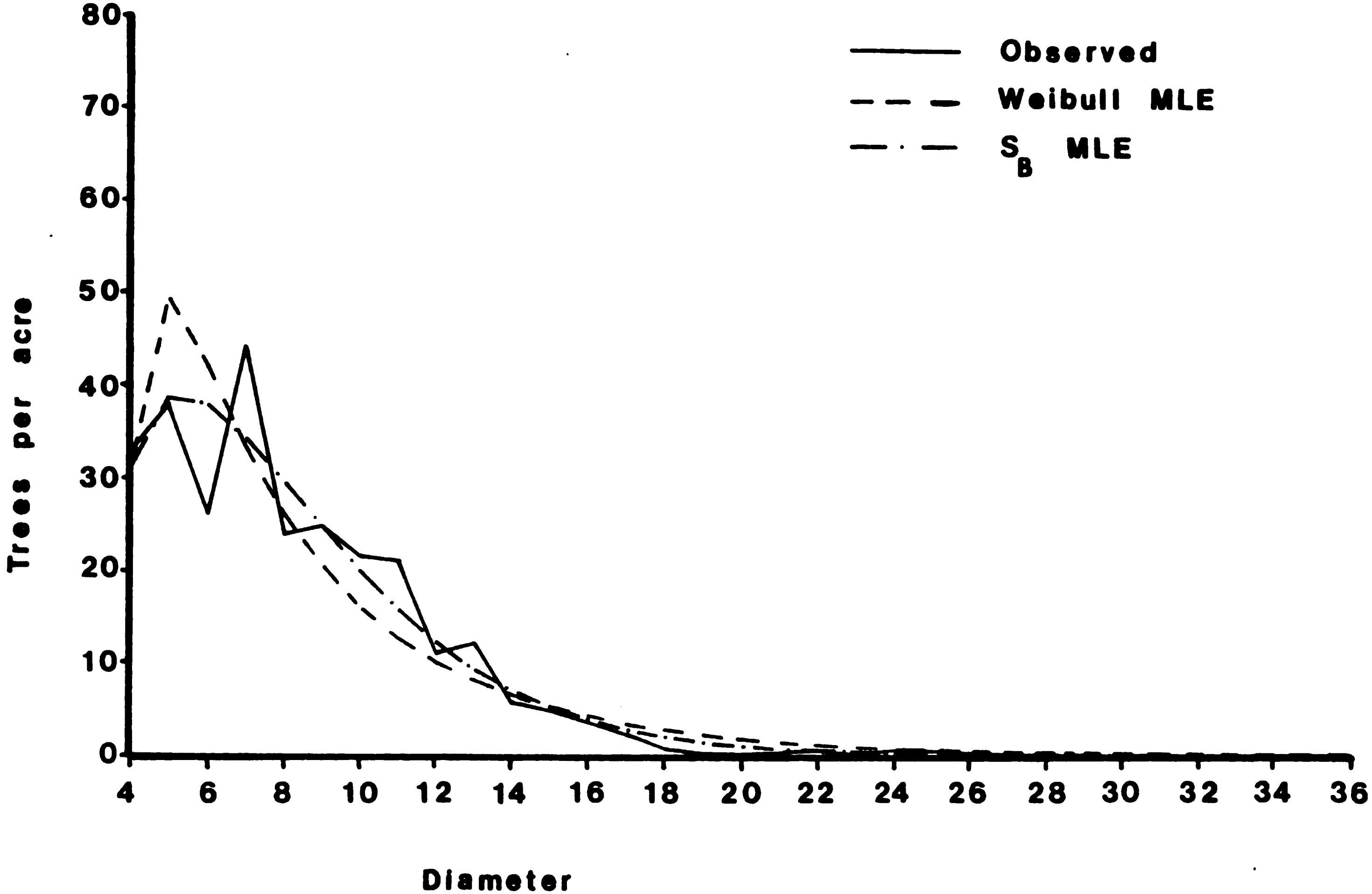

Figure 21: Observed and predicted (maximum likelihood estimation) diameter distributions for ELTP 45.

Summary and Conclusions

In general, the nature of the data analyzed in this study presents a complex modelling problem. First, the ELTP's are composed of mixed upland hardwood species and, to some extent, mixed age groups. This sort of forest as a whole has not yet been considered in diameter distribution modelling in the literature. Second, the ELTP's are composed of stands with disparate diameter distributions. The expected and actual result of these two conditions is a series of somewhat regular to very irregular diameter distributions.

The use of point sample data is suitable for modelling ELTP diameter distributions. The methods given are based on sampling and distribution theory taken from sampling with probability proportional to size. A larger number of sample points per stand may lead to improved accuracy of small diameter class (< 8") frequencies. Overall, the MLE models especially provide good fits to the observed data except in cases where the observed distributions are irregular to very irregular (ELTP's 35 and 37). ELTP 37 may present a case where the compound distribution modelling method of Cao and Burkhart (1984) would be appropriate. Otherwise, the good fits to observed data indicate that point sample data are useful in modelling observed diameter distributions. This point is especially important since a great deal of forest inventory is done with point sampling.

The grouping of frequency data before estimating parameters may result in diminished accuracy. However, the MLE models in particular provide results that are sufficiently good to state that the possible loss of accuracy may not be great enough to offset the ease of the method. The use of the grouped frequency counts should include an acknowledgement of this possible shortcoming.

The fact that any of the models used in this study produce a good overall fit to the observed data may be taken as evidence that these methods are worth consideration. In these methods, three new facets of tree diameter distribution modelling are given: modelling of mixed species and mixed age upland hardwood diameter distributions; the use of grouped frequency counts for distribution parameter estimation; and the use of point sample data for modelling diameter distributions. Using the grouped frequencies facilitates the use of a microcomputer and spreadsheet software for calculations. The chief advantages in this are that a mainframe computer is not necessary and manipulation of data is reduced.

As to the models themselves, the Weibull distribution provides the best overall accuracy with these data with the least effort. The S_B distribution is theoretically more flexible and therefore more interesting than the Weibull, but its estimation accuracy is not commensurate with its greater complexity. Therefore, the Weibull distribution is preferred in this study.

The relatively poorer performance of the S_B with respect to the Weibull may be related to the parameter estimation methods chosen for the S_B. The simplest and most explicit methods available are described and used in this study but these methods are not as simple or explicit as those for the Weibull distribution. This weakness may be the cause of the unacceptable level of accuracy of the S_B as compared to the Weibull. In addition, the S_B may be more sensitive than the Weibull to the use of grouped frequencies for parameter estimation. The result of inaccurate parameter estimates is an inaccurate fit of predicted to observed data. These speculations serve to point out that the most significant impediment to the common use of the S_B is the difficulty in parameter estimation previously described.

Both parameter estimation methods for the Weibull distribution are explicit and straight forward. The percentile estimation method as given in this study is the easier of the two methods for parameter estimation. The Weibull PCTE model is not as accurate in predicting class frequencies as the Weibull MLE model, but for estimation of a somewhat generalized nature, the PCTE model is appropriate. If more accuracy is required, the PCTE estimates of c are a good starting point for the iterative solution of the maximum likelihood estimate of c. The Weibull MLE parameter estimates are suitable for inferences about stand structure.

ELTP-level parameters are predicted most accurately from the MLE models, particularly the Weibull. However, the increase in accuracy is marred by a slight bias in prediction. The PCTE model predictions yield unbiased results, except for predictions of basal area per acre. It should be noted, however, that the 95% confidence intervals about the errors are wider for the PCTE models than for the MLE models. This wider confidence interval indicates an overall lower reliability in individual predictions. Based on the extremes of the error confidence intervals and consideration of the very small size of the biasedness in most biased predictions, both methods produce acceptable accuracy for all ELTP-level parameters except for the PCTE model predictions of basal area per acre. Acceptable accuracy in these cases is defined as error less than 10% of the observed parameter mean across ELTP's.

The ability of the Weibull model to accurately predict ELTP-level parameters means that it provides an accurate summary of the ELTP's. An accurate summary is desirable for growth and yield modelling. One feature of growth and yield modelling based on a distribution model is that it allows diameter class specificity without the necessity of modelling individual trees (Little, 1983). A distribution model that is accurate in class frequencies and summary at the outset will be less subject to error through projection cycles.

The Weibull PCTE model is most accurate in predicting observed distribution skewness and kurtosis. In doing so, this model does best in modelling observed curve shape as measured by skewness and kurtosis. However, the difference in accuracy between the Weibull and S_B models in predicting skewness and kurtosis is not great except that the S_B models produce biased estimates of skewness. In general, prediction of observed skewness and kurtosis gives a much more generalized look at curve fitting than the goodness-of-fit tests. The results of this study suggest that the Weibull model and its parameter estimation methods provide unbiased estimates of observed curve shape.

In general, the use of spreadsheet software for estimating distribution parameters and obtaining distribution frequencies provides a quick and easy alternative to mainframe computing. This method may bring diameter distribution modelling within the reach of those who are interested in diameter distribution modelling but who don't have access to more expensive and powerful hardware systems.

To summarize, this study examines the relative performance of the Weibull distribution and the S_B distribution in modelling ELTP diameter distributions. The ELTP's are composed of upland hardwoods of somewhat mixed age groups. Conventional distribution parameter estimation methods are described and applied to grouped frequency counts of trees per acre by diameter class as obtained from point sample data.

The Weibull distribution is preferred for modelling the observed ELTP diameter distributions. The Weibull PCTE model provides an adequate approximation to observed data on its own. In addition, the Weibull PCTE model provides a good starting point for maximum likelihood parameter estimation for the Weibull distribution.

APPENDICES

Appendix I.

Ecological Land Type Phase Descriptions

The following ELTP descriptions are from Cleland, Hart, Pregitzer, Host, and Padley (March, 1986; unpublished). The descriptions are not in any way complete but are intended to provide general classification background as may be desired for further understanding of the diameter distributions developed in this study. The classifications are considered unofficial at this time since the study is still in progress. Therefore, ELTP as presented in this study is not a definitive classification.

Species names:

Black oak : Quercus velutina

White oak : Quercus alba

(Northern) red oak : Quercus rubra

(Upland) pin oak : Quercus ellipsoidalis

Red maple : Acer rubrum

Sugar maple : Acer saccharum

White ash : Fraxinus americana

(Bigtooth and quaking) aspen : Populus grandidentata
Populus tremuloides

Basswood : Tilia americana

Beech : Fagus grandifolia

ELTP 1: Pin oak-white oak-Deschampsia plant association on excessively well drained sands of outwash plains. Overstory composition is upland pin oak and white oak. The canopy is relatively open and ground flora coverage is very low. Red maple does not occur in the understory.

Stand averages for the overstory:

species	BA/a	Gross Vol	SI
pin oak	53	1033	56
white oak	24	325	49
	81	1438	

age = 72 MAI = 20.6 cu.ft./a/yr.

ELTP 10: Black oak-white oak-Vaccinium plant association on excessively well drained sands of outwash plains. Similar to ELTP 1 except in having a more closed canopy and in the presence of red maple and bracken fern in the understory. Overstory composition is black oak, white oak, and northern red oak.

species	BA/a	Gross vol	SI
black oak	32	767	50
white oak	32	606	42
red oak	18	388	54
	85	1817	

age = 81 MAI = 23.1 cu.ft./a/yr.

ELTP 12: Black oak-white oak-Vaccinium plant association on sub-irrigated, excessively well drained sands of outwash plains. Similar to ELTP 10 except for the presence of sub-irrigation and the wider presence of witch hazel as a shrub. Sub-irrigation refers to the presence of a water table within tree rooting depth for an extended period of time. Overstory composition is the same as ELTP 10.

species	BA/a	Gross vol	SI
black oak	35	828	56
white oak	40	685	49
red oak	7	178	54
	90	1832	

age = 73 MAI = 25.3 cu.ft./a/yr.

ELTP 20: Mixed oak-red maple-low Viburnum plant association on well to excessively well drained sands on overwashed moraines, kame terraces, spillways, and outwash plains. The overstory of this ELTP is made up of northern red oak, white oak, black oak, and red maple. Red maple and witch hazel are well represented in the 1-3" classes.

species	BA/a	Gross vol	SI
red oak	30	819	61
white oak	30	689	52
black oak	35	865	60
red maple	5	71	
	102	2533	

age = 81 MAI = 31.2 cu.ft./a/yr.

ELTP 21: Mixed oak-red maple-low Viburnum plant association. Similar to ELTP 20 except for the absence of sub-irrigation and the presence of sandy loam textural bands beneath well to somewhat excessively well drained sands. Overstory composition is northern red oak, white oak, black oak, and red maple.

species	BA/a	Gross vol	SI
red oak	40	967	65
white oak	30	682	53
black oak	21	534	66
red maple	11	144	
	107	2460	

age = 72 MAI = 34.4 cu.ft./a/yr.

ELTP 35: Red oak-red maple-high Viburnum plant association on well drained sands with fine loamy substrata on moraines and overwashed lake beds. Overstory composition is northern red oak, white oak, and red maple.

species	BA/a	Gross vol	SI
red oak	73	2166	77
white oak	14	372	63
red maple	14	274	74
	117	3169	

age = 72 MAI = 44.3 cu.ft./a/yr.

ELTP 37: Red oak-red maple-Desmodium plant association on well to moderately well drained sandy loams over loamy substrata on ground moraines and fine textured glacial lakebeds. Overstory composition is northern red oak, white oak, and red maple.

species	BA/a	Gross vol	SI
red oak	62	2192	85
white oak	20	498	63
red maple	25	615	69
	113	3561	

age = 73   MAI = 48.9 cu.ft./a/yr.

ELTP 40: Sugar maple-beech-Lycopodium plant association on well drained morainal sands. Overstory composition is sugar maple, beech, northern red oak, and red maple.

species	BA/a	Gross vol	SI
sugar maple	27	589	65
red oak	32	973	76
red maple	7	66	65
	104	2717	

age = 62   MAI = 43.6 cu.ft./a/yr.

ELTP 43: Sugar maple-northern red oak-Lycopodium plant association on well drained morainal sands with fine textured substrata. Overstory composition is sugar maple, beech, northern red oak, and red maple.

species	BA/a	Gross vol	SI
sugar maple	40	969	73
red oak	50	1840	88
red maple	2	60	73
	119	3649	

age = 67 MAI = 54.3 cu.ft./a/yr.

ELTP 45: Sugar maple-white ash-Osmorhiza plant association on well to moderately well drained morainal sands over fine substrata. Overstory composition is sugar maple, white ash, northern red oak, and red maple.

species	BA/a	Gross vol	SI
sugar maple	52	1388	76
red oak	9	343	86
white ash	14	471	85
	126	3868	

age = 66 MAI = 59.5 cu.ft./a/yr.

Appendix II

S_B Percentile Parameter Estimation Tables

The tables in this appendix may be used to facilitate estimation of the parameters of the S_B distribution by percentiles. Use of these tables is mentioned in the materials and methods section (Chapter 3).

The tables are based on the percentile estimation methods of Shapiro and Slifker (1980). These methods are described in Chapter 3. The sample size (n) appears in the first column on the left. The second column contains the values of z, the tabulated standard normal variate (Steele and Torrie, 1980). The remaining four columns contain the four symmetric percentiles according to the sample size and z-value.

Given a sample size and an initial value of z, the corresponding percentiles are read from the table. These percentiles identify individual observations from a list of ordered observations (in this case, diameter). The four appropriate observations are then used to calculate m, n, and p. The parameter estimates can then be obtained accordingly via the hyperbolic trigonometric functions given in Chapter 3.

The sample sizes include 50 and 75 and otherwise are given by tens from 100 through 400. Interpolation of percentiles between sample sizes is possible. Likewise, interpolation between z-values is possible.

Table 21:
Tables for percentile estimation
of Johnson's distribution parameters

n	z	1	2	3	4
50	0.5	3.84	15.93	35.08	47.16
50	0.6	2.30	14.22	36.79	48.71
50	0.7	1.40	12.60	38.40	49.61
50	0.8	0.91	11.10	39.91	50.09
50	0.9	0.68	9.71	41.30	50.33
75	0.5	5.51	23.64	52.36	70.49
75	0.6	3.19	21.07	54.93	72.81
75	0.7	1.84	18.65	57.35	74.16
75	0.8	1.12	16.39	59.61	74.89
75	0.9	0.76	14.31	61.69	75.24
100	0.5	7.18	31.35	69.65	93.82
100	0.6	4.09	27.93	73.07	96.91
100	0.7	2.29	24.70	76.30	98.71
100	0.8	1.32	21.69	79.31	99.68
100	0.9	0.85	18.91	82.09	100.15
110	0.5	7.85	34.44	76.57	103.15
110	0.6	4.45	30.67	80.33	106.55
110	0.7	2.47	27.12	83.88	108.53
110	0.8	1.40	23.81	87.19	109.60
110	0.9	0.89	20.75	90.25	110.12
120	0.5	8.52	37.52	83.48	112.48
120	0.6	4.81	33.42	87.58	116.19
120	0.7	2.65	29.54	91.46	118.35
120	0.8	1.48	25.93	95.07	119.52
120	0.9	0.92	22.59	98.41	120.08
130	0.5	9.18	40.61	90.40	121.82
130	0.6	5.17	36.16	94.84	125.83
130	0.7	2.83	31.96	99.04	128.17
130	0.8	1.57	28.05	102.95	129.43
130	0.9	0.96	24.43	106.57	130.05
140	0.5	9.85	43.69	97.31	131.15
140	0.6	5.53	38.90	102.10	135.47
140	0.7	3.01	34.38	106.62	137.99
140	0.8	1.65	30.17	110.83	139.35
140	0.9	0.99	26.27	114.73	140.01

n	z	1	2	3	4
150	0.5	10.52	46.78	104.23	140.48
150	0.6	5.89	41.65	109.36	145.11
150	0.7	3.19	36.80	114.20	147.82
150	0.8	1.73	32.29	118.72	149.27
150	0.9	1.03	28.12	122.89	149.98
160	0.5	11.19	49.86	111.14	149.81
160	0.6	6.24	44.39	116.61	154.76
160	0.7	3.36	39.22	121.78	157.64
160	0.8	1.81	34.40	126.60	159.19
160	0.9	1.06	29.96	131.04	159.94
170	0.5	11.86	52.95	118.06	159.14
170	0.6	6.60	47.13	123.87	164.40
170	0.7	3.54	41.64	129.36	167.46
170	0.8	1.89	36.52	134.48	169.11
170	0.9	1.10	31.80	139.20	169.91
180	0.5	12.52	56.03	124.97	168.48
180	0.6	6.96	49.87	131.13	174.04
180	0.7	3.72	44.06	136.94	177.28
180	0.8	1.98	38.64	142.36	179.02
180	0.9	1.13	33.64	147.36	179.87
190	0.5	13.19	59.12	131.89	177.81
190	0.6	7.32	52.62	138.38	183.68
190	0.7	3.90	46.48	144.52	187.10
190	0.8	2.06	40.76	150.24	188.94
190	0.9	1.17	35.48	155.52	189.84
200	0.5	13.86	62.20	138.80	187.14
200	0.6	7.68	55.36	145.64	193.32
200	0.7	4.08	48.90	152.10	196.92
200	0.8	2.14	42.88	158.12	198.86
200	0.9	1.20	37.32	163.68	199.80
210	0.5	14.53	65.29	145.72	196.47
210	0.6	8.04	58.10	152.90	202.96
210	0.7	4.26	51.32	159.68	206.74
210	0.8	2.22	45.00	166.00	208.78
210	0.9	1.24	39.16	171.84	209.77
220	0.5	15.20	68.37	152.63	205.80
220	0.6	8.40	60.85	160.15	212.60
220	0.7	4.44	53.74	167.26	216.56
220	0.8	2.30	47.12	173.88	218.70
220	0.9	1.27	41.00	180.00	219.73

n	z	1	2	3	4
230	0.5	15.86	71.46	159.55	215.14
230	0.6	8.76	63.59	167.41	222.24
230	0.7	4.62	56.16	174.84	226.38
230	0.8	2.39	49.24	181.76	228.61
230	0.9	1.31	42.84	188.16	229.70
240	0.5	16.53	74.54	166.46	224.47
240	0.6	9.12	66.33	174.67	231.88
240	0.7	4.80	58.58	182.42	236.20
240	0.8	2.47	51.36	189.64	238.53
240	0.9	1.34	44.68	196.32	239.66
250	0.5	17.20	77.63	173.38	233.80
250	0.6	9.48	69.08	181.93	241.53
250	0.7	4.98	61.00	190.00	246.03
250	0.8	2.55	53.48	197.53	248.45
250	0.9	1.38	46.53	204.48	249.63
260	0.5	17.87	80.71	180.29	243.13
260	0.6	9.83	71.82	189.18	251.17
260	0.7	5.15	63.42	197.58	255.85
260	0.8	2.63	55.59	205.41	258.37
260	0.9	1.41	48.37	212.63	259.59
270	0.5	18.54	83.80	187.21	252.46
270	0.6	10.19	74.56	196.44	260.81
270	0.7	5.33	65.84	205.16	265.67
270	0.8	2.71	57.71	213.29	268.29
270	0.9	1.45	50.21	220.79	269.56
280	0.5	19.20	86.88	194.12	261.80
280	0.6	10.55	77.30	203.70	270.45
280	0.7	5.51	68.26	212.74	275.49
280	0.8	2.80	59.83	221.17	278.20
280	0.9	1.48	52.05	228.95	279.52
290	0.5	19.87	89.97	201.04	271.13
290	0.6	10.91	80.05	210.95	280.09
290	0.7	5.69	70.68	220.32	285.31
290	0.8	2.88	61.95	229.05	288.12
290	0.9	1.52	53.89	237.11	289.49
300	0.5	20.54	93.05	207.95	280.46
300	0.6	11.27	82.79	218.21	289.73
300	0.7	5.87	73.10	227.90	295.13
300	0.8	2.96	64.07	236.93	298.04
300	0.9	1.55	55.73	245.27	299.45

n	z	1	2	3	4
310	0.5	21.21	96.14	214.87	289.79
310	0.6	11.63	85.53	225.47	299.37
310	0.7	6.05	75.52	235.48	304.95
310	0.8	3.04	66.19	244.81	307.96
310	0.9	1.59	57.57	253.43	309.42
320	0.5	21.88	99.22	221.78	299.12
320	0.6	11.99	88.28	232.72	309.01
320	0.7	6.23	77.94	243.06	314.77
320	0.8	3.12	68.31	252.69	317.88
320	0.9	1.62	59.41	261.59	319.38
330	0.5	22.54	102.31	228.70	308.46
330	0.6	12.35	91.02	239.98	318.65
330	0.7	6.41	80.36	250.64	324.59
330	0.8	3.21	70.43	260.57	327.79
330	0.9	1.66	61.25	269.75	329.35
340	0.5	23.21	105.39	235.61	317.79
340	0.6	12.71	93.76	247.24	328.29
340	0.7	6.59	82.78	258.22	334.41
340	0.8	3.29	72.55	268.45	337.71
340	0.9	1.69	63.09	277.91	339.31
350	0.5	23.88	108.48	242.53	327.12
350	0.6	13.07	96.51	254.50	337.94
350	0.7	6.77	85.20	265.80	344.24
350	0.8	3.37	74.67	276.34	347.63
350	0.9	1.73	64.94	286.07	349.28
360	0.5	24.55	111.56	249.44	336.45
360	0.6	13.42	99.25	261.75	347.58
360	0.7	6.94	87.62	273.38	354.06
360	0.8	3.45	76.78	284.22	357.55
360	0.9	1.76	66.78	294.22	359.24
370	0.5	25.22	114.65	256.36	345.78
370	0.6	13.78	101.99	269.01	357.22
370	0.7	7.12	90.04	280.96	363.88
370	0.8	3.53	78.90	292.10	367.47
370	0.9	1.80	68.62	302.38	369.21
380	0.5	25.88	117.73	263.27	355.12
380	0.6	14.14	104.73	276.27	366.86
380	0.7	7.30	92.46	288.54	373.70
380	0.8	3.62	81.02	299.98	377.38
380	0.9	1.83	70.46	310.54	379.17

n	z	1	2	3	4
390	0.5	26.55	120.82	270.19	364.45
390	0.6	14.50	107.48	283.52	376.50
390	0.7	7.48	94.88	296.12	383.52
390	0.8	3.70	83.14	307.86	387.30
390	0.9	1.87	72.30	318.70	389.14
400	0.5	27.22	123.90	277.10	373.78
400	0.6	14.86	110.22	290.78	386.14
400	0.7	7.66	97.30	303.70	393.34
400	0.8	3.78	85.26	315.74	397.22
400	0.9	1.90	74.14	326.86	399.10

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Bailey, Robert L. 1980. Individual tree growth derived from diameter distribution models. Forest Science 26:626-632.

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

3. Bowman, K.O., C.A. Serbin, and L.R. Shenton. 1981. Explicit approximate solutions for S_B. Communications in Statistics-Sim. and Comp. 10:1-15.

4. Cao, Quang V. and Harold E. Burkhardt. 1984. A segmented approach for modelling diameter frequency data. Forest Science 30:129-137.

5. Cohen, Clifford A. 1965. Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. Technometrics 7:579-588.

6. Conover, W.J. 1980. Practical nonparameteric statistics. 2nd ed. John Wiley and Sons, New York. 491 pp.

7. Dubey, Satya D. 1967. Some percentile estimators for Weibull parameters. Technometrics 9:119-129.

8. Dubey, Satya D. 1967. On some permissable estimators of the location parameter of the Weibull and certain other distributions. Technometrics 9:293-307.

9. Ek, Alan R., James N. Issos, and Robert L. Bailey. 1975. Solving for Weibull diameter distribution parameters to obtain specified mean diameters. Forest Science 21:290-292.

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

11. Husch, B., C.I. Miller, and T.W. Beers. 1982. Forest mensuration. 2nd ed. John Wiley and Sons, New York. 402 pp.

12. Hyink, David, M. and John W. Moser, Jr. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science 29:85-95.

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

14. Johnson, N.L. and Samual Kotz. 1970. Continuous univariate distributions - I. 300pp. Houghton Mifflin Co., Boston.

15. Little, Susan N. 1983. Weibull diameter distributions for mixed stands of western conifers. Canadian Journal of Forest Research 13:85-88.

16. Lorimer, Craig G. and Albert G. Krug. 1983. Diameter distributions in even-aged stands of shade-tolerant to midtolerant tree species. The American Midland Naturalist 109:331-345.

17. Lorimer, Craig G. and Lee E. Frelich. 1984. A simulation of equilibrium diameter distributions of sugar maple (Acer saccharum). Bulletin of the Torrey Botanical Club 3:193-199.

18. Mage, David T. 1980. An explicit solution for S_B parameters using four percentile points. Technometrics 22:247-251.

19. Rauscher, H.M. 1986. The microcomputer scientific software series 4: Testing prediction accuracy. Submitted for publication.

20. Slifker, James F. and Samuel s. Shapiro. 1980. The Johnson system: selection and parameter estimation. Technometrics 22:239-246.

21. Steele, R.G.D. and J.H. Torrie. 1980. Principles and procedures of statistics. 2nd ed. McGraw-Hill, New York. 633 pp.

22. Weibull, W. 1951. A statitical distribution function of wide applicability. Journal of Applied Mechanics 18:293-297.

23. Zarnoch, Stanley J. and Tommy R. Dell. 1985. An evaluation of percentile and maximum likelihood estimators of Weibull parameters. Forest Science 31:260-268.

24. Zarnoch, S.J., C.W. Ramm, V.J. Rudolph, and M.W. Day. 1980. The effects of red pine thinning regimes on diameter distribution fitted to the Weibull function. MSU Ag Exp Sta Res Report 423. 10pp.