

Models of Stand Basal Area Distributions, Individual Tree  
Basal Area Growth,  
and Height-Diameter Relationships for Loblolly Pine

by

Edwin James Green

Dissertation submitted to the Graduate Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Forestry

APPROVED:

-----  
H. E. Burkhardt, Chairman

-----  
R. G. Oderwald

-----  
M. R. Reynolds

-----  
J. E. Mann

-----  
D. Wm. Smith

July, 1981  
Blacksburg, Virginia

## ACKNOWLEDGEMENTS

First and foremost I must thank Dr. H.E. Burkhart. His support and guidance throughout the term of this investigation was invaluable. I am sure I can never repay him for the many hours he willingly sat and listened to my impulsive ideas. I must also thank the members of the reading committee, Drs. R.G. Oderwald, J.E. Mann, M.R. Reynolds, and D.Wm. Smith, for their time and thoughtful review of the work reported herein. I would like to express my gratitude to Dr. T.E. Burk, who substituted for Dr. Oderwald during my oral examination. Thanks also are due

who was most helpful during my brief excursion into the field of stochastic differential equations. I am indebted to and

for providing me with their computer programs to compute A.P.A. and fit the Weibull distribution, respectively.

I am grateful to the School of Forestry and Wildlife Resources and the Loblolly Pine Growth and Yield Cooperative of VPI and SU for making this study possible, both physically and financially. I am also indebted to the Louisiana State University Hill Farm Experiment Station and the Weyerhaeuser Company for the data used in this study.

Of course it almost goes without saying that I must thank my fellow graduate students for their suggestions and criticisms. Especially helpful among this group were

, , and

I would like to thank my parents for their wisdom and guidance throughout my life. Finally, I must thank my wife,

, who typed some of the tables in this document and put up with me for the last three long years, and my sons,

and , who provided many welcome respites from the rigors of graduate work.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	ii
Chapter	page
1. INTRODUCTION . . . . .	1
Background for study . . . . .	1
Objectives . . . . .	2
2. REVIEW OF THE LITERATURE . . . . .	4
Individual tree models . . . . .	4
Diameter-distribution models . . . . .	6
Diameter and basal area growth . . . . .	7
Height-diameter/basal area curves . . . . .	10
Chapman-Richards generalized growth function . . . . .	13
Stochastic growth functions . . . . .	20
3. DATA . . . . .	25
4. METHODS . . . . .	28
Height-diameter/basal area curves . . . . .	28
Parameter Distribution model . . . . .	33
Specifying A and c a priori . . . . .	35
Sensitivity analysis for A . . . . .	36
Estimation of k and m . . . . .	37
Characterizing the marginal distribution of k . . . . .	41
Prediction of m . . . . .	46
Survival function . . . . .	47
Application . . . . .	48
Nonlinear Growth Function . . . . .	51
Specification of model . . . . .	51
Application . . . . .	57
5. RESULTS AND DISCUSSION . . . . .	60
Height-diameter/basal area curve . . . . .	60
Choice of curve . . . . .	60
Choice of surrogate regressions . . . . .	63
Parameter Distribution model . . . . .	66
Fitting of growth curves . . . . .	66
Estimating the marginal distribution of k . . . . .	69
Prediction of m from k . . . . .	76
Survival curve . . . . .	77

Application to independent data . . . . .	77
Nonlinear growth function . . . . .	100
6. CONCLUSIONS AND RECOMMENDATIONS . . . . .	104
LITERATURE CITED . . . . .	109
7. VITA . . . . .	116

## LIST OF FIGURES

Figure	page
1. Several common mathematical models for describing biological growth . . . . .	16
2. Observed basal area distribution at age 21 on plot 41 of the Hill Farm data, and basal area distribution predicted with the parameter distribution model . . . . .	84
3. Observed basal area distribution at age 21 on plot 42 of the Hill Farm data, and basal area distribution predicted with the parameter distribution model . . . . .	85
4. Observed basal area distribution at age 20 on plot 43 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model . . . . .	86
5. Observed basal area distribution at age 20 on plot 44 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model . . . . .	87
6. Observed basal area distribution at age 21 on plot 45 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model . . . . .	88
7. Observed height distribution at age 21 on plot 41 of the Hill Farm data, and height distributions predicted with the parameter distribution model . . . . .	89
8. Observed height distribution at age 21 on plot 42 of the Hill Farm data, and height distributions predicted with the parameter distribution model . . . . .	90
9. Observed height distribution at age 20 on plot 43 of the Hill Farm data, and height distributions predicted with the parameter distribution model . . . . .	91
10. Observed height distribution at age 20 on plot 44 of the Hill Farm data, and height distributions predicted with the parameter distribution model . . . . .	92
11. Observed height distribution at age 21 on plot 45 of the Hill Farm data, and height distributions predicted with the parameter distribution model . . . . .	93

LIST OF TABLES

Table	page
1. Common height-diameter (basal area) curves . . . . .	11
2. Several common mathematical models for describing biological growth . . . . .	17
3. Observed basal area at age 20 and predicted basal areas at ages 20 and 50 with three different asymptotes for four individual loblolly pine trees from each thinning treatment on replication 3 in the Hill Farm data . . . . .	38
4. Mean estimated values and percentage of cases in which zero was not included in a 95% confidence interval for the shape parameter $m$ of the Chapman-Richards growth curve, by plot for replication three of the Hill Farm data . . . . .	40
5. Probability density functions (p.d.f.'s) for distributions selected as candidates for the marginal distribution of $k$ , the rate parameter of the Chapman-Richards growth curve . . . . .	43
6. Mean estimated crown ratios for loblolly pine trees, predicted with the equation by Feduccia et al. (1979), by plot and age for replication one of the Hill Farm data . . . . .	55
7. Summed ranks over 105 plot-age combinations on the Hill Farm data for five candidate height-diameter (basal area) models, based on three statistics . . . . .	61
8. Statistics from fitting the candidate surrogate regressions for $a$ and $b$ in the model $H = a + b \log(\text{BA})$ on the first three replications of the Hill Farm data . . . . .	65
9. Mean and standard deviation of $R^2$ values from fitting the Chapman-Richards growth curve to all loblolly pines on a plot in the Hill Farm data, by plot . . . . .	68
10. Values and ranks for $D = \max(D^+,  D^- )$ from fitting the five candidate distributions for the marginal distribution of $k$ , the rate parameter of the Chapman-Richards growth curve, to estimates of $k$ resulting from fitting the curve to individual loblolly pine trees on three replications in the Hill Farm data . . . . .	71
11. Frequency of ranks within plots for $D = \max(D^+,  D^- )$ from fitting the five candidate distributions for the marginal distribution of $k$ , the rate parameter of the Chapman-Richards growth curve, to estimates of $k$ resulting from fitting	

	the curve to individual loblolly pine trees on three replications in the Hill Farm data . . . . .	72
12.	Probabilities from one sample K-S test of the fit of the Weibull distribution to estimates of k, the rate parameter of the Chapman-Richards curve, by plot for three replications in the Hill Farm data . . . . .	73
13.	Statistics for the candidate models for prediction of the two parameters, b and c, of the Weibull distribution used to characterize the marginal distribution of k, the rate parameter of the Chapman-Richards growth curve. Models fit on, and k values estimated from, three replications in the Hill Farm data . . . . .	75
14.	Means and standard deviations of basal area (sq. ft.) observed on one replication of the Hill Farm data and of predicted basal area, and p-levels from comparison of observed and predicted basal area distributions via the two-sample K-S test . . . . .	80
15.	Means and standard deviations of height (feet) observed on one replication in the Hill Farm data and of predicted height, and p-levels from comparison of predicted and observed height distributions via the two-sample K-S test . . . . .	81
16.	Mean and standard deviation of basal area (sq. ft.) observed on one replication of the Hill Farm data, and of means and standard deviations of predicted basal area based on 20 runs of the parameter distribution model . . . . .	82
17.	Mean and standard deviation of height (feet) observed on one replication in the Hill Farm data, and of means and standard deviations of predicted height based on 20 runs of the parameter distribution model . . . . .	83
18.	Means and standard errors of means of difference and absolute difference between basal area observed on one replication in the Hill Farm data, and basal area predicted with the linear model and with the nonlinear model with three different asymptotes . . . . .	103

## Chapter 1

### INTRODUCTION

#### 1.1 BACKGROUND FOR STUDY

The objective of growth and yield studies is to develop models to predict product volumes which may be expected for given sets of stand characteristics. Management alternatives may then be devised to effect those characteristics which are expected to result in favorable volume production. Whole-stand models have been developed which predict some specified aggregate volume of a stand for a given set of stand characteristics. However, these models are not sufficient for many situations. Different size standards for different products often necessitate the prediction of volumes of given portions of a stand, or of sets of trees of specified sizes. To meet these needs, models have been developed to predict either the size (diameter and/or height) distribution within a stand or the sizes of individual trees in the stand for given sets of stand characteristics. Logically, the growth components in these models should mimic the underlying growth process as closely as possible. Additionally it is sometimes desirable to include a stochastic element in growth models. However, in nearly all growth and yield models the mechanism for

generating diameter or basal area growth is either explicitly or implicitly empirically constructed. Additionally, in size-distribution models, often no allowance is made for a stochastic element. Well-defined growth functions, from a biological standpoint, provide some assurance that the growth equations in individual tree size or size distribution models have desirable properties. An individual tree size model developed along these lines can be made stochastic by attaching a random variable to the predicted size, as is conventionally done. This is valid because of the large number of individual tree sizes ordinarily predicted. In size distribution models, where only a limited number of size classes are projected through time, this method is not practical and an alternative method of introducing random variability into these models is needed.

Priority should be placed on modeling diameter or basal area growth because this is the single most important tree dimension from a utilization standpoint. Heights may then be predicted from diameters using one of the common height-diameter curves.

## 1.2 OBJECTIVES

The objectives of this study were:

1. To develop an individual tree growth function, based on the Chapman-Richards curve, in which the parameters of the curve are modelled as functions of tree and stand variables, and apply the function to basal area growth of loblolly pine trees.
2. To compare the results of the model developed under objective (1) to those obtained from a simpler linear model of basal area growth on an independent data set.
3. To develop and apply a stochastic model to predict basal area distributions, based on the the Chapman-Richards curve, in which the parameters of the curve are considered to be dependent random variables.
4. To study various height-diameter curves in order to determine which is the most appropriate for the data set utilized, and apply this curve as part of the model developed in accord with objective (2).

## Chapter 2

### REVIEW OF THE LITERATURE

#### 2.1 INDIVIDUAL TREE MODELS

Models which describe the growth of individual trees have generally been divided into two categories: distance-independent and distance-dependent (Munro, 1974). In the former, the spatial coordinates of the trees are not required, and the measure of competition included does not utilize inter-tree distances. In these models tree growth is simulated either on an individual stem basis, or by size classes. Thus some of these models actually yield only size distributions, while others yield individual tree sizes. Stochastic variation has been included in size-class distribution models through a Markov Chain approach (see Section 2.6). A stochastic element may be incorporated into models which predict individual tree sizes, whether distance-dependent or distance-independent, by adding a random variable to predicted tree sizes (see Section 2.6).

No uniform structure has been adhered to in the development of distance-independent models, so it is difficult to make any general statements concerning their content (Daniels and Burkhart, 1975). Examples of distance-independent models may be found in Goulding (1972), Stage (1973), Dale (1975), or Bruner and Moser (1973).

Although the detail varies from model to model, distance-dependent models developed in the past have followed a similar overall structure. In these models, individual stem growth in height and diameter is predicted via functions of site quality and the stem's current size and competitive status. The competition measure is generally a function of the stem's size in relation to the sizes of and distances to its neighbors. Thus the spatial location of each stem is required. Distance-dependent models generally provide more detailed information about individual tree growth than distance-independent models. However, distance-dependent models are also more expensive to develop and employ and have more stringent data requirements than distance-independent models. Examples of distance-dependent models may be found in Newnham and Smith (1964), Bella (1970), Hatch (1971), Arney (1972), Ek and Monserud (1974), Mitchell (1975), Daniels and Burkhardt (1975), and Daniels et al. (1979). Daniels and Burkhardt (1975) provide a thorough review of the literature pertinent to distance-dependent models.

## 2.2 DIAMETER-DISTRIBUTION MODELS

Diameter distribution models represent a major class of growth and yield models for predicting size distributions. These models are based upon characterization of the diameter distribution present in a stand at a given point in time with a probability density function (p.d.f.). Typically, a p.d.f. is selected and fit to the diameter distributions observed in sample data. The estimates of the parameters of the distribution are then regressed upon stand-level variables (e.g. site index, number of trees per acre, age). The diameter distribution may then be predicted for any given set of stand characteristics. Height-diameter curves are ordinarily used to estimate the mean total height of trees of given diameter classes growing under given conditions. With estimates of the total number of trees present, the relative frequency of each diameter class (obtained from the predicted p.d.f.), and mean total height per diameter class, it is possible to estimate volume for any portion of the stand. Diameter distribution models do not ordinarily include a stochastic component.

These models differ chiefly in the p.d.f. used. When the method was first developed, the beta distribution was predominantly used (Clutter and Bennett, 1965). More recently the Weibull distribution has been used (Bailey and

Dell, 1973). Some limited use has been made of the Johnson's  $S_B$  distribution (Hafley and Schreuder, 1977). The primary difficulty in the implementation of these models, regardless of p.d.f., has been prediction of the parameters of the density function. The prediction models employed have generally fit the observed data poorly. To circumvent this problem a new technique of predicting quantities which are functions of the moments of the distributions (e.g. quadratic mean diameter) and then solving back for the parameters has recently been developed (Matney and Sullivan, in press; Hyink, in press) This approach is still too new for any general conclusions to be drawn as to its utility.

The major diameter distribution models developed for loblolly pine (Pinus taeda, L.) thus far may be found in Lenhart and Clutter (1971), Lenhart (1972), Smalley and Bailey (1974), Burkhart and Strub (1974), and Feduccia et al. (1979).

### 2.3 DIAMETER AND BASAL AREA GROWTH

To date, individual tree diameter or basal area growth functions have generally been empirical in nature. One common approach (Daniels and Burkhart, 1975 or Stage, 1973) has been to derive an empirical regression equation

describing diameter growth of open-grown trees. This quantity is then reduced according to an index of competition. This methodology is based on the premise that for a given age and site quality, maximum diameter increment is obtained in the absence of competition. It is a long established fact that competition in the form of increased density results in reduced diameter increment. In fact, this is the basis for the common practice of thinning.

Others (Lohrey, 1974; Jackson and Ure, 1964; Alemdag, 1978; Newnham and Mucha, 1971) have developed empirical linear models to predict increment directly, using a measure of competition as an independent variable. Increment has also been predicted directly via nonlinear equations (Alder, 1979; Hahn and Leary, 1979; Daniels, 1981). Still others have developed linear regressions to predict the logarithm of diameter increment, adjusted for climatic factors (Errikson, 1978).

Bailey (1980) proposed another type of diameter increment model. He derived the increment equation implied by use of growth and yield models based on diameter distributions. As previously discussed, in these models the current diameter distribution is characterized by a probability density function (p.d.f.), e.g., the beta or Weibull, and then the parameters of the distribution are

projected through time. Thus, once an initial p.d.f. is selected, the diameter distribution is forced to conform to that p.d.f. for the duration of the projection. The maintenance of a given p.d.f. implies a transformation of diameters through time, and Bailey identified the transformations and resulting implicit growth functions which correspond to most of the commonly used p.d.f.'s. For example, Bailey showed that use of the Weibull p.d.f. implies a growth function of the form

$$W = A_0 e^{bt^k}$$

where: W = size  
 t = time  
 e = base of natural logarithms  
 b, k = arbitrary constants  
 A<sub>0</sub> = initial size (at time zero).

This is a specific case of a generalized growth function presented by Prodan (1968, pp. 370); however, Prodan did not investigate the properties of this curve. The growth rate implied by this model is given by

$$\frac{dW}{dt} = A_0 e^{bt^k} bkt^{k-1} = Wbkt^{k-1}$$

Obviously, b and k must be of the same sign in order for increment to be positive. If b and k are negative, the curve decreases from time zero, reaches a minimum, and then increases and is asymptotic at A<sub>0</sub>, the initial size. This is obviously undesirable from a biological point of view.

Bailey applied the model and computed estimates for  $b$  and  $k$  which were both positive, a condition which implies uninterrupted exponential growth.

#### 2.4 HEIGHT-DIAMETER/BASAL AREA CURVES

Foresters have long known that there is correlation between diameter (or basal area) and height of individual stems in even-aged stands. Thus it is common to predict mean total tree height from diameter or basal area. Numerous curve forms have been reported, and the most common are presented along with their apparent authors in Table 1. The origin of model 2 is unclear. Models 6 and 7 are modifications of models 5 and 4 respectively, designed to take into account that a tree has no d.b.h. (measured at 4.5 feet above ground) until it reaches 4.5 feet in height.

Curtis (1967) conducted a comprehensive study of the performance of many different height-diameter/basal area curves in second-growth Douglas-fir (*Pseudotsuga menziesii*, (Mirb.) Franco). He ranked the various models based on Furnival's index of fit (Furnival, 1961) and found that with the exception of models 5 and 6, all of the models presented in Table 1 performed well. Based primarily upon ease of application, he recommended the use of models 3 and 4. However, he cautioned against the use of model 3 with data

Table 1. Common height-diameter (basal area) curves.

Curve*	Author(s)
1. $H = a + b D + c D^2$	Trorey 1932
2. $H = a + b (1/D) + c (1/D^2)$	----
3. $H = a + b \log(D)$	Henricksen 1950
4. $\log(H) = a + b (1/D)$	Michailoff 1943
5. $\log(H) = a + b \log(D)$	Stoffels and Van Soest 1953
6. $\log(H-4.5) = a + b \log(D)$	----
7. $\log(H-4.5) = a + b (1/D)$	----
8. $H = 4.5 + h (1 - e^{-aD})$	Meyer 1940

\* H = total tree height.

D = dbh.

e = base of natural logarithms.

a, b, c = constants to be estimated which vary from curve to curve.

h = asymptote of (H-4.5).

from young ages, since it might predict negative heights for small basal areas.

Ek (1973) studied the performance of most of the common height-diameter/basal area curves in cases where the sample sizes were small (i.e., less than or equal to 10). From a population of 600 trees, random and systematic samples were drawn of size 3, 5, and 10. For each sample size and selection procedure, coefficients for each of the height equations were computed. Using these coefficients, total height was predicted for every tree in the population. The height models were then evaluated on the basis of bias and precision. Ek concluded that models 3 and 4 were among the better height predictors for small sample sizes, but that model 3 had a tendency to predict negative heights for small stems. Models 5 and 6 were not included in Ek's study.

Model 5 is based on the allometric relationship

$$H = aD^b$$

and has been used by Stage (1975), Stout (1973), and Greenhill (1881) (as reported by Stout (1973)). The allometric relationship between height and diameter was studied by Stage (1975) and Pienaar and Turnbull (1973) and determined to be satisfactory for Western U.S. coniferous species and Norway spruce (Picea abies, (L.) Karst.), respectively. Stout (1973) recommended that this model be

modified to take into account that the origin should be 4.5 feet in height if used with data from young trees.

Model 4 has been widely applied in studies of southern pines (Burkhart and Strub, 1974; Matney and Sullivan, in press; Daniels, 1981). Lenhart (1972), Lenhart and Clutter (1971), Smalley and Bailey (1974) and Feduccia et al. (1979) have all used modifications of this model. The reasons for this model's popularity appear to be four-fold: first, it is easy to apply; second, it yields an upper asymptote for height; third, it cannot predict negative heights; and fourth, it has performed adequately in the aforementioned studies. Note that regarding the first reason, all the models in Table 1 are easy to apply. Regarding the second, if diameter and height are predicted together by a system of equations and if the diameter model is asymptotic over time, then implicitly all the height models in Table 1 are asymptotic over time.

Model 8, presented by Meyer(1940) is essentially the monomolecular growth function, with the dependent variable being height above 4.5 feet and the independent variable being diameter.

## 2.5 CHAPMAN-RICHARDS GENERALIZED GROWTH FUNCTION

As stated in the introduction, the component growth functions in growth and yield models should mimic biological growth as closely as possible. This may be accomplished by using a biological growth function as a base, and constructing the tree growth function upon this base. In the last century or so, many mathematical functions have been developed and employed for the purpose of describing the growth of an organism. Richards (1969) performed extensive analyses of most of the common functions. Prodan (1968), Turnbull (1963) and Grosenbaugh (1965) have studied many of the same functions as they relate to forestry applications. These functions differ primarily in the shapes they may assume. Several of the more common functions are shown in Figure 1 and Table 2. The two extreme cases are the exponential and the monomolecular. The exponential function has no inflection point or upper asymptote, and implies a growth rate which is proportional to current size and a linearly increasing function through time. The monomolecular function also lacks an inflection point. In this function, the growth rate is a constant proportion of growth yet to be made, i.e., the difference between the asymptote and current size. The growth rate is thus a linearly decreasing function through time. The remaining growth curves all fall somewhere between the

exponential and monomolecular forms. Some functions, such as the logistic or Backman, have an inflection point at size  $A/2$ , where  $A$  is the upper asymptote, and are symmetric about this point (Prodan, 1968; Richards, 1959). Other functions, such as the Gompertz (Richards, 1959) or Von Bertalanffy (Von Bertalanffy, 1957), have inflection points below the point  $A/2$  and are asymmetric.

Von Bertalanffy's growth function was based upon the allometric relationship:

$$S = bV^m$$

where  $S$  represents surface area,  $V$  is volume or weight, and  $b$  and  $m$  are constants. Since the cube root of volume has a linear dimension, the square of the cube root has the same dimension as surface area. Thus  $m$  in the above relationship should have a value of  $2/3$ . Von Bertalanffy (1957) hypothesized that anabolism (constructive metabolism) is proportional to surface area, while catabolism (destructive metabolism) is proportional to volume. Potential volume growth rate was then defined as the difference between anabolism and catabolism, and was expressed as:

$$\text{Potential Volume Growth Rate} = aV^m - bV.$$

Studies of various aquatic and terrestrial organisms led Von Bertalanffy to conclude that volume growth of many organisms conformed to this equation, while for others the value of  $m$

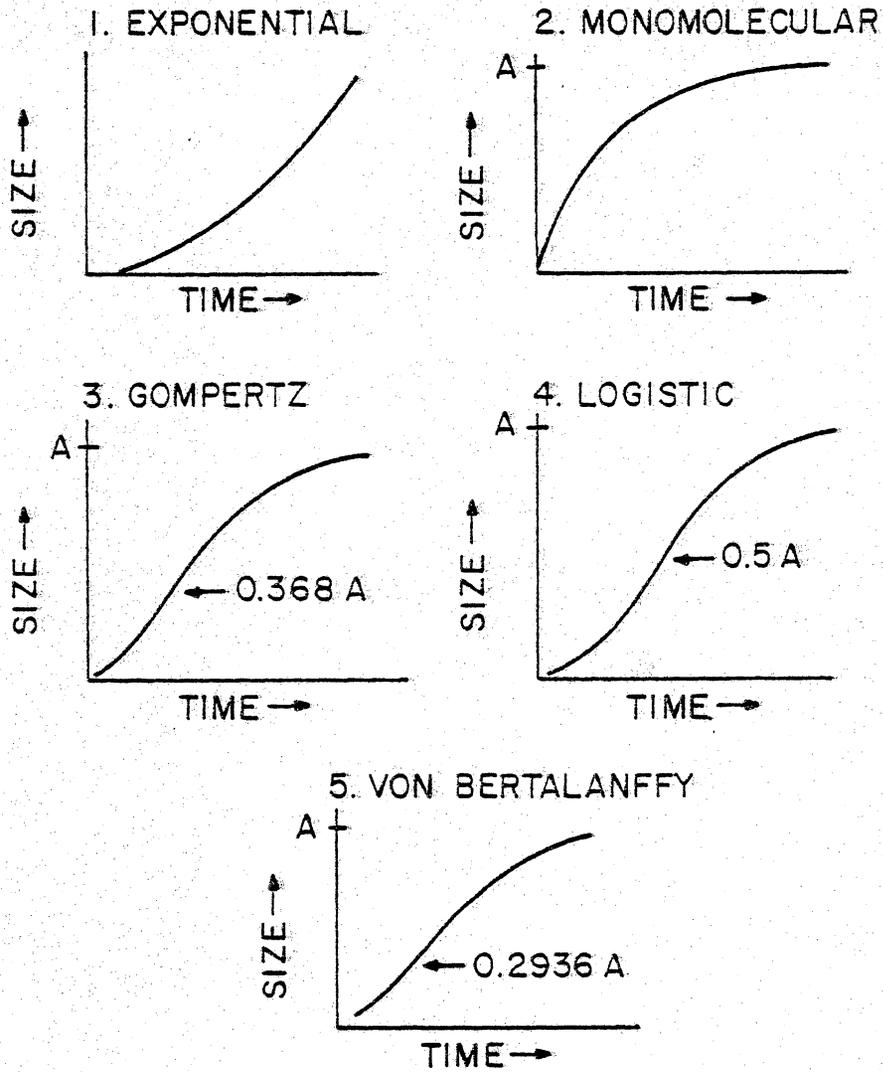


Figure 1. Several common mathematical models for describing biological growth ( $A$  = asymptote; arrow marks location of inflection point. Adapted from Turnbull, 1963).

Table 2. Several common mathematical models for describing biological growth.\*

Model	Cumulative growth function	Increment function**
exponential	$W = W_0 e^{kt}$	$w = kW$
monomolecular	$W = A(1 - be^{-kt})$	$w = k(A - W)$
Von Bertalanffy	$W = \left[\frac{n}{k}\right]^{1/3} - \left[\frac{n}{k}\right]^{1/3} e^{-kt/3}$	$w = nW^{2/3} - kW$
Gompertz	$W = Ae^{-be^{-kt}}$	$w = kW \log(A/W)$
Logistic	$W = \frac{A}{1 + be^{-kt}}$	$w = \frac{k(A - W)W}{A}$
Chapman-Richards	$W^{1-m} = A^{1-m}(1 - be^{-kt})^{\#}$	$w = \frac{kW\left(\frac{A}{W}\right)^{1-m} - 1}{1-m}$

\* W = cumulative size,

t = time.

A = asymptotic final size,

e = base of natural logarithms.

log = natural logarithm.

$W_0, k, b, m, n$  = constants which vary from model to model.

\*\*  $w = dW/dt$ ,

# if  $m > 1, b < 0$ ; if  $m < 1, b > 0$ .

varied between  $2/3$  and  $1$ . When  $m=2/3$ , the cumulative growth curve is asymmetrical and sigmoid, with the inflection point occurring at about  $1/3$  of final volume (see Figure 1). When  $m=1$ , the growth rate is a constant proportion of current size and growth is exponential.

Chapman (1961) and Richards (1959,1969) both studied the Von-Bertalanffy curve and suggested it be generalized through reparameterization to the following form:

$$W = A(1 - be^{-kt})^{1/(1-m)}$$

$$\begin{array}{l} \text{if } m < 1, b > 0 \\ m > 1, b < 0 \end{array}$$

where  $W$  is size,  $A$  is the upper asymptote,  $t$  is time,  $e$  is the base of the natural logarithms, and  $b$ ,  $k$ , and  $m$  are constant parameters. Richards (1959) showed that in this form  $m$  was really a shape parameter. By allowing  $m$  to assume any positive value, Richards showed that the curve was able to take on a great many forms. In particular, when  $m=0$ , the curve reduces to the monomolecular function; when  $m=2/3$ , the Von Bertalanffy; when  $m=1$ , the Gompertz (actually when  $m=1$  the differential equation has no solution, but Richards (1959) has shown that the solution approaches the Gompertz curve as  $m$  approaches  $1$ ); when  $m=2$ , the logistic; and as  $m$  approaches infinity, the exponential. Thus as  $m$  increases, the inflection point moves outward from the

origin ( $m=0$ ) to infinity ( $m$  infinite). The remaining parameters of the curve are  $A$ ,  $k$ , and  $b$ .  $A$  is the asymptote or final size. Parameter  $b$  reflects the choice of origin and is biologically meaningless (Richards, 1959). The final parameter,  $k$ , is a rate parameter. It measures the rate of increase of some quantity, but what this quantity is is conditional upon the value of  $m$  (Richards, 1959). This generalized curve has become known by various names, among them the Chapman-Richards (C-R) curve, the Richards curve, and the generalized Von Bertalanffy curve. Following the publication of the C-R curve, few new growth curves have been developed.

The C-R model has been applied in studies of forest stand development by Turnbull (1963), Pienaar (1965), Moser (1967), Pienaar and Turnbull (1973) and Rose and Chen (1977). Pienaar and Turnbull (1973) also demonstrated the adequacy of the C-R curve for describing individual tree growth in volume, basal area, and height for Norway spruce trees grown in Austria. The data used by Pienaar and Turnbull (1973) were taken from Assman (1970), who took it from Guttenburg (1915). Neither Assman (1970) nor Pienaar and Turnbull (1973) describe the conditions of the stands in which the stems were grown. However, since Guttman's (1915) study included stems over 100 years of age, based upon the

silvics of the species and the history of European forestry, it can be assumed the stands were natural and even-aged. Finally, the C-R curve has been used to describe height growth in connection with site index studies (Beck, 1971; Carmean, 1972; Graney and Burkhart, 1973; Burkhart and Tennent, 1977).

## 2.6 STOCHASTIC GROWTH FUNCTIONS

It has long been recognized that a deterministic growth methodology is not completely appropriate for an organism situated in a dynamic environment. Forest scientists have attempted to model unaccounted for variability by incorporating a random element into growth predictions. Three different methods have been employed to accomplish this, and at least one other, as yet untried, method exists.

The conventional method of generating a random growth increment or random cumulative size (depending on whether the investigator uses a differential or integral equation) in models which predict individual tree sizes has been to develop an equation to predict the mean growth of an individual stem possessing a given set of characteristics. A normally distributed random deviate is then added to the predicted quantity. The variance of the random component is taken to be equal to the Mean Square Error (MSE) resulting

from the fitting of the growth function to observed data (Daniels and Burkhardt, 1975). This approach has generally been found to perform satisfactorily, and is in agreement with ordinary regression theory. In this theory it is usually assumed that the conditional p.d.f. of the dependent variable is normal, that the conditional variance is homogeneous and well estimated by the MSE, and that the model is correct (thus the random deviate added to predicted growth has mean zero).

Another method for incorporating stochastic variability into growth predictions has been utilization of Markov Chain theory (Buongiorno and Michie, 1980; Peden et al., 1973; Bruner and Moser, 1973). This approach has been applied exclusively in distance-independent models describing growth in uneven-aged stands. In general, the probability of a tree in the  $i$ -th state (diameter class) moving up 0 or more states, dying, being harvested, or achieving some other specified condition during a specified time interval is estimated from an existing data set. These probabilities are then applied across an existing diameter distribution in the form of a transition matrix, thus determining the predicted distribution at the end of the interval. Bruner and Moser (1973) provide an excellent description of the mechanics of the method. This approach

seems to work well for short term predictions, but as the length of the time interval is increased the accuracy of this method declines (Bruner and Moser, 1973). Two crucial assumptions are routinely made in this approach. One is the stationarity assumption which requires the transition probabilities to be static. The other is the Markov assumption which requires that the system have no memory, i.e., future development depends only on the present state of the system. Both of these assumptions have been challenged on biological grounds (Bruner and Moser, 1973).

The third and most recent method of stochastic growth prediction has been the use of stochastic differential equations (García, 1979). A stochastic differential equation is basically an ordinary differential equation with a random component added on. The solution to a stochastic differential equation is not a point estimate but rather a distribution. In order to make the solution mathematically tractable, one must ordinarily make two assumptions. First one must assume that the basic underlying model is linear. Second, one ordinarily assumes that the random component is a Weiner process. The relevant property of this process is that successive increments are assumed to be independent and normally distributed (Hoel et al., 1972). In a growth context, the underlying model is generally believed to be

non-linear, and successive increments are known to be highly correlated. Furthermore, since increments cannot be negative, they cannot be truly normally distributed.

García (1979) adopted the C-R growth curve to stochastically model height development. He made a transformation to make the equation linear, and he accepted the Weiner process. With this methodology he developed site index curves for radiata pine (Pinus radiata, D. Don) that compared favorably with those developed by Burkhart and Tennent (1977). However, Burkhart and Tennent used a much simpler methodology.

A fourth method of stochastic growth prediction exists, although it has not yet been attempted in forestry. In this method the parameters of the growth curve are considered to be random variables. In the case of linear regression, a sizeable quantity of literature exists on estimating stochastic parameters (Rao, 1965). These methods generally assume a known error structure. Unfortunately, in the case of the growth of individual organisms, the models are generally nonlinear and the error structure is generally unknown. Krause et al. (1967) utilized a three parameter logistic growth function in a study on the growth of chickens. They assumed the parameters to be stochastic and independent. They were then able to solve for the joint

distribution of the parameters. However, Karish (1973) has shown that when the parameters of a nonlinear growth function are estimated by nonlinear least squares, a high degree of correlation among the estimates is generally found. Thus the independence assumption of Krause et al. (1967) is suspect. Little other work along these lines has been reported (Goldstein, 1979).

## Chapter 3

### DATA

Data used in this study came from the Mutual Competition Study at the North Louisiana Hill Farm Experiment Station, Homer, Louisiana. The entire tract on which the experiment is located is classified as site index 68, base age 25 years for loblolly pine. However, the height data indicate that site quality is not quite so uniform. The soils are predominantly fine sandy loams. Prior to implementation of the experiment the area was an abandoned crop field. Before planting, the vegetation on the area was cut and the stumps burned.

The area was planted with loblolly pine in 6 x 6 foot spacing in February, 1958. Twenty 0.25-acre plots were established and segregated into four blocks or replications based on height in 1962. However, no significant differences in mean height could be detected among the four replications. In 1962, four plots (one from each replication) were thinned to 1000 TPA (trees per acre). The remaining 16 plots were thinned to 600 TPA in the same year. When a .1 inch difference in mean annual diameter growth was detected between the plots thinned to 1000 TPA and those thinned to 600 TPA, 12 (four on each replication) plots were

thinned to 300 TPA. The same procedure was followed for the treatments of 200 and 100 TPA. The process was completed in 1965. Hereafter, the plots will be referred to by a two digit number. The first number will identify the replication and the second the treatment. The treatment code is:

1. 1000 TPA
2. 600 TPA
3. 300 TPA
4. 200 TPA
5. 100 TPA.

Thus, for example, plot 23 is the plot on the second replication which was thinned to 300 TPA.

The thinnings were combination selection-spacing thinnings, designed to leave high quality residual stems in as uniform a spacing pattern as possible. It may be expected that this resulted in higher quality residual stands at the lower densities than at the higher densities. This is a confounding factor which should be taken into consideration when comparing growth responses between plots thinned to different residual densities.

Measurements were made in nine years: 1962, 1963, 1964, 1965, 1966, 1967, 1970, 1977, and 1978. Whenever a plot was measured, the diameter of every tree on the plot was

recorded to the nearest .01 inch. Exceptions to the measurement schedule were as follows: Plot 11 was not measured in 1964 or 1978. Plots 21, 31, and 41 were not measured in 1964. Plots 15, 32, 43, and 44 were not measured in 1978. Thus, although the maximum age from planting in the data set is 21 years, the maximum age at which all plots were measured is 20 years. Heights of every tree on each plot were measured to the nearest .1 foot in 1962 and 1963. Heights were also measured for every tree in 1978 if the plot was measured that year, for every tree on plots thinned to 300 or fewer TPA in 1977, and for every tree on plots thinned to 100 TPA each year except 1964 (and 1978 for plot 15). Whenever a plot was measured and not all heights were recorded, a sample of trees was measured for heights. The samples were subjectively selected so as to be representative of diameter distribution on the plot. No well-defined sampling procedure was followed.

Finally, the spatial coordinates of every tree left after completion of the thinning treatments were recorded in 1977 so it is possible to construct individual tree competition indices based on inter-tree distances and relative heights.

## Chapter 4

### METHODS

#### 4.1 HEIGHT-DIAMETER/BASAL AREA CURVES

As mentioned in the Literature Review there are many published functions relating total height of an individual stem to its diameter or basal area. Most of them seem to perform fairly well (Curtis, 1967), and thus there does not appear to be a need to search for a new, probably more complex function. It appears that time would be better spent identifying which of the existing models is the most appropriate for a given data set. Consequently, five of the more common height-diameter/basal area curves were studied to determine which performed best on the Hill Farm data. The 'best' model was then incorporated into the Parameter Distribution model (see next section).

The five models selected for comparison were:

1.  $H = a + b \log(BA)$
2.  $\log(H) = a + b \log(BA)$
3.  $\log(H-4.5) = a + b \log(BA)$
4.  $\log(H) = a + b (1/D)$
5.  $\log(H-4.5) = a + b (1/D)$

where H is total height, D is diameter at breast height (d.b.h.), BA is basal area, a and b are constants which vary from model to model, and the logarithms are to base e.

Models 1 and 4 are those suggested for use by Curtis (1967) whereas 2 is the allometric model used by Stage (1975) and Stout (1973). Models 3 and 5 are modifications of 2 and 4 designed to take into account the fact that individual trees have no d.b.h. or basal area until they reach 4.5 feet in height.

The models were compared on the first three replications within the data set, and the fourth replication was reserved for applying the height models within the Parameter Distribution model. On the first three replications, each model was fit by ordinary least squares to all the height-diameter/basal area pairs on each plot at each measurement. This resulted in a total of 105 plot-age combinations. For each combination the five models were ranked in terms of squared multiple correlation coefficient ( $R^2$ ), PRESS, and SUMABS. PRESS, an acronym for prediction sum of squares, gives an indication of a model's predictive ability. It is calculated by removing one of the  $n$  observations at a time from the data set and calculating the regression coefficients based on the  $(n-1)$  remaining observations. Values of the independent variables for the removed observation are inserted into the calculated regression and an estimate of the dependent variable is computed. Differences between the estimate and the observed value are then squared and summed over  $n$ , i.e.,

$$\text{PRESS} = \sum (Y_i - \hat{Y}_i)^2$$

where:  $Y_i$  = the i-th observation on the dependent variable

$\hat{Y}_i$  = the estimator of  $E(Y_i)$  excluding the i-th observation

$n$  = the number of observations in the data set.

The model(s) which yield the lowest PRESS value(s) should be those which would predict the best on an independent data set. SUMABS is defined here as the sum of the absolute PRESS residuals, i.e.,

$$\text{SUMABS} = \sum |Y_i - \hat{Y}_i|$$

where the variables remain as defined above. Thus SUMABS is another useful prediction oriented tool. SUMABS is not constrained to give the same results as PRESS, so both statistics were used. For the models which predict  $\log(H)$  or  $\log(H-4.5)$  the three statistics ( $R^2$ , PRESS, and SUMABS) had to be transformed since height is the variable of interest, not its logarithm. The value of  $R^2$  was determined for the  $\log(H)$  models as follows:

$$R^2 = \{ \sum (\hat{H}_i - \bar{H})^2 / \sum (H_i - \bar{H})^2 \}$$

where:  $H_i$  = total height of tree i.

$\bar{H}$  = mean total tree height

$\hat{H}_i$  = anti-log of the predicted value of the dependent variable for the ith observation on the independent variable, i.e.,

$\hat{H}_i = \exp(\hat{Y}_i)$  where  $Y_i = \log(H_i)$

For  $\log(H-4.5)$  models  $R^2$  was computed in an analogous manner, except that 4.5 feet was added to the predicted value after taking anti-logs.

The method applied in computing PRESS and SUMABS for the logarithmic models was basically the same as that for computing  $R^2$ . After the  $i$ -th observation was removed from the data set, and an estimate for the  $i$ -th independent variable computed from the  $(n-1)$  remaining observations, the resulting value was converted to height and the difference between it and the actual observed height was computed. PRESS was then the sum of the squared differences, and SUMABS the sum of the absolute differences.

Once the five models were ranked in terms of  $R^2$ , PRESS, and SUMABS for each plot-age combination the ranks were summed over all 105 combinations, and the 'best' model chosen based on the summed ranks.

Using a procedure followed by Burkhart and Strub (1974) and Matney and Sullivan (in press), surrogate regressions were then determined for the coefficients  $a$  and  $b$  of the selected model. These surrogate regressions involved only stand-level variables. The variables tested for inclusion in each regression were:

1. quadratic mean diameter
2. basal area per acre

3. thinning treatment
4. age since thinning treatment
5. age since planting
6. average height of dominants and codominants
7. log of average height of dominants and codominants
8. number of trees per acre
9. log of number of trees per acre

All logarithms were to the base  $e$ . Values of  $R^2$  were determined for every possible linear combination of the preceding variables for each dependent variable (a and b). The models which performed best in terms of  $R^2$  were then tested using PRESS. Then after consideration of  $R^2$ , PRESS, SUMABS, and the sum of the PRESS residuals (SUMRES), a surrogate regression was selected for each of the two variables a and b. These were inserted into the overall height model. This was then fit to all the height-diameter/basal area pairs on the first three replications at every measurement time simultaneously, as the surrogates were designed to take into account changing stand characteristics. The final fitted height model was then implemented within the Parameter Distribution model, as described in the next section.

#### 4.2 PARAMETER DISTRIBUTION MODEL

The fourth method of stochastic growth projection mentioned in the Literature Review, that of considering the parameters of the growth curve to be random variables, was used in constructing a model to predict basal area distributions. This model is called the Parameter Distribution model. Because of its great flexibility, the C-R curve was adopted as the basic growth function. The following form of the function was used:

$$BA = A(1 - e^{-k(t-c)})^{1/(1-m)} \quad (1)$$

where: BA = basal area

A = asymptotic BA

t = time

k, c, m = constant parameters

e = base of natural logarithms

In this form the curve is conditioned such that BA=0 when t=c. Thus c represents the time required for a seedling to reach 4.5 feet in height. A in (1) is the asymptotic basal area of the individual tree. The remaining two parameters, k and m, are the rate and shape parameters of the C-R curve respectively.

The purpose of the Parameter Distribution model was to produce estimated basal area and height distributions for

given sets of stand characteristics. The data was split into a fitting set and a validation set. The C-R curve was fit to every stem in the fitting data set. The parameters A and c were fixed a priori, for reasons described below, leaving two free parameters, k and m. For each plot, a sampling distribution was then fit to the estimates of k, and models developed to predict the parameters of the sampling distribution from stand variables. A model was then developed to predict m from k and stand variables. Finally, an existing survival curve was modified for use in this model. To operate the model, the number of trees present at a given point in time is predicted from the survival curve. Next the parameters of the sampling distribution for k are predicted. A k value is then randomly selected from this distribution for each stem predicted to be present on the basis of the survival curve. For each k value, an m value is then predicted. With these two values, the predicted basal area of the tree is then computed from the C-R curve. Repeating this process for each stem predicted to be present results in a predicted basal area distribution. Additionally, a height is predicted for each stem using the fitted height-basal area

curve, resulting in a predicted height distribution.

#### 4.2.1 Specifying A and c a priori

In order to simplify the estimation process, A and c were fixed a priori. A was fixed because the data set used included ages only up to 21 years, and it would have been difficult to estimate the asymptotic final basal area. The parameter c was fixed because the time required for individual stems to reach 4.5 feet in height on similar sites (such as those in the Hill Farm data set) should be essentially constant. Pienaar and Turnbull (1973) made a similar assumption.

The value for A was chosen by searching the literature for the largest recorded basal area for a loblolly pine. Based on the American Forestry Association's register of big trees (Pardo, 1973) the largest basal area for loblolly pine is approximately 21.9 square feet. On the assumption that the asymptotic basal area should be slightly larger than the largest recorded, this figure was increased to 22.3 square feet (approx. 64 in. d.b.h.). It may be argued that many trees would never reach this asymptote, but would die first. A counter-argument may be that these stems are killed by factors external to the growth processes of the stems, and that if the trees were to live forever, they would

eventually achieve this asymptotic size. At any rate, fixing A simplified the estimation process considerably.

The value of  $c$  was chosen by randomly selecting 15 trees from the data set and fitting equation (1) to them with A fixed, but  $c$ ,  $k$ , and  $m$  estimated from the data by nonlinear least squares. The nonlinear least squares procedure used at this point and for the remainder of this study was that of Marquardt. This procedure generally performs better than alternative procedures when the parameters are highly correlated (SAS Institute, 1979), as they generally are in growth curves (Karish, 1973).

Estimates for  $c$  varied from .7 to 6.3, with 3 being the approximate median. It was felt that 3 years following planting represented a reasonable estimate for  $c$ , and  $c$  was fixed at 3 for the remainder of the study.

#### 4.2.2 Sensitivity analysis for A

Since the choice of A was arbitrary, it was the opinion of the author that the sensitivity of the C-R curve to A should be investigated. For this purpose, A was fixed at 18.96 sq. ft. (15% below 22.3), 22.3 sq. ft., and 25.65 sq. ft. (15% above 22.3) and  $k$  and  $m$  were estimated ( $c=3$  yrs.) for four trees randomly selected from each of the five thinning treatments on the third replication. Using these

estimates, the basal areas of the trees were computed at age 20 and 50 for each of the three values of  $A$ . These sizes are shown, along with the observed basal area at age 20 in Table 3. The values for the three asymptotes are similar, suggesting that the curve is relatively insensitive to  $A$ . For each value of  $A$ , the curve seemed to slightly over-predict the basal area at age 20 for the high densities (1000 and 600 TPA). However, in only one case was the over-prediction greater than 10% of the observed value. It was concluded that 22.3 sq. ft. was an acceptable value for  $A$ , and this value was employed for the remainder of the study.

The value of  $c$  was considered to be reasonable, therefore no sensitivity analysis was performed for this parameter.

#### 4.2.3 Estimation of $k$ and $m$

The next step was to fit equation (1) to individual trees on the first three replications of the data set (the fourth was reserved for validation purposes). Only trees with at least three measurements were included. However, preliminary fittings of equation (1) revealed a tendency, especially at the high densities, for  $m$  to either be estimated at 0 or for 0 to be included in the asymptotic 95% confidence intervals for  $m$ . In order to test whether  $m$

Table 3. Observed basal area at age 20 and predicted basal areas at ages 20 and 50 with three different asymptotes for four individual loblolly pine trees from each thinning treatment on replication 3 in the Hill Farm data.\*

Thinning treatment (TPA)	Observed BA (ft <sup>2</sup> ) at age 20	Predicted					
		Age 20 asymptote			Age 50 asymptote		
		18.96	22.3	25.65	18.96	22.3	25.65
1000	0.1650	0.1725	0.1725	0.1725	0.4730	0.4736	0.4740
1000	0.1722	0.1850	0.1850	0.1850	0.5071	0.5078	0.5083
1000	0.4128	0.4270	0.4271	0.4271	1.1573	1.1609	1.1636
1000	0.1051	0.1352	0.1352	0.1352	0.3714	0.3718	0.3721
600	0.3150	0.3434	0.3435	0.3436	0.9343	0.9368	0.9387
600	0.1475	0.1617	0.1617	0.1617	0.4436	0.4442	0.4446
600	0.3234	0.3356	0.3357	0.3357	0.9134	0.9158	0.9176
600	0.3490	0.3858	0.3859	0.3860	1.0476	1.0509	1.0531
300	0.4516	0.4687	0.4688	0.4688	1.2678	1.2722	1.2754
300	0.6480	0.6383	0.6383	0.6383	1.8015	1.8078	1.8124
300	0.4922	0.4854	0.4854	0.4854	1.5284	1.5341	1.5384
300	0.5454	0.5741	0.5742	0.5742	1.5451	1.5516	1.5564
200	0.4516	0.4465	0.4465	0.4465	1.4932	1.4994	1.5042
200	0.3068	0.3069	0.3069	0.3069	0.8364	0.8382	0.8396
200	0.5786	0.5667	0.5667	0.5666	1.6876	1.6936	1.6980
200	0.7854	0.7762	0.7761	0.7761	2.6954	2.7178	2.7349
100	1.1946	1.1783	1.1781	1.1780	4.1718	4.2288	4.2722
100	1.1309	1.1053	1.1051	1.1049	4.2196	4.2838	4.3332
100	1.2767	1.2838	1.2837	1.2836	3.5926	3.6198	3.6398
100	0.7594	0.7725	0.7725	0.7724	2.2180	2.2278	2.2351

\* Basal area in square feet.

might be set equal to 0, a preliminary fitting was performed using all the stems on replication 3 only. The results of this fitting are shown in Table 4. Since the percent of cases where 0 was not included in the asymptotic 95% confidence intervals for m was less than the conventional Type I error rate of .05 for plots 31 and 32, it was decided to to set m to 0 for densities 1000 and 600 TPA. When m is fixed, equation (1) may be transformed to

$$\log[1 - (BA/22.3)^{(1-m)}] = -k(t-c) \quad (2)$$

with k the unknown parameter. Equation (2) is a no-intercept linear model and may be fit by ordinary least squares. Thus equation (2) was fit to individual trees on plots 11, 12, 21, 22, 31, and 32, while equation (1) was fit by nonlinear least squares to all the stems on plots 13, 14, 15, 23, 24, 25, 33, 34, and 35. As with equation (1), only trees with at least three measurements were fit with equation (2).

A justification for setting m, the shape parameter of the C-R curve, to 0 for the denser plots may be made on biological grounds. As m increases, the position of the inflection point moves upward from the origin (i.e., no inflection point) to infinity. When m=0, the growth rate is maximum at time zero and declines over time. When m>0, the

Table 4. Mean estimated values and percentage of cases in which zero was not included in a 95% confidence interval for the shape parameter  $m$  of the Chapman-Richards growth curve, by plot for replication three of the Hill Farm data.

---

---

Plot	n	Percent cases	Mean $m$
31	247	4.88	0.0151
32	150	4.67	0.0114
33	73	16.00	0.0763
34	50	66.00	0.1650
35	24	75.00	0.2778

---

growth rate increases from time zero until the the time at which the inflection point occurs, and then declines. It may be argued that in less dense stands a state of juvenile growth exists during which the individual tree grows at an increasing rate. This stage continues until competition (crown or root or both) or some other ecological phenomenon sets in, and then the growth rate begins to remain constant or decline. In denser stands this juvenile state either does not exist or terminates so early that it is of no practical consequence. This logic appears to be borne out by the means of the estimated  $m$  values in Table 4. A trend is evident with the means increasing with decreasing density. Only the means of plots 31 and 32 violate this trend. It appears that the density at which this hypothetical juvenile growth rate is prolonged into a range of practical interest occurs somewhere between 600 and 300 TPA.

#### 4.2.4 Characterizing the marginal distribution of $k$

Following the fitting of equations (1) and (2) to the stems on the first three replications, the marginal distribution of  $k$  for each plot was fit by a p.d.f. Since  $k$  must be positive the choice of candidate distributions was limited to those allowing only positive values. The three

candidates selected were the lognormal, Weibull, and  $S_B$  distributions. The p.d.f.'s of these three distributions are shown in Table 5. The lognormal was chosen because many biological phenomena tend to follow a lognormal distribution (Krause et al., 1967). The Weibull and  $S_B$  were chosen because of their familiarity to foresters. This familiarity stems from the use of these two distributions in diameter distribution growth and yield models (Bailey and Dell, 1973; Hafley and Schreuder, 1977).

Each of the three distributions was fit to the estimated  $k$  values on each plot using the method of maximum likelihood to estimate the distributional parameters. Maximum likelihood estimators (MLE's) may be shown to have many desirable statistical properties (Bickel and Doksum, 1977).

The MLE's for the lognormal distribution exist in closed form. The MLE's for the Weibull distribution must be found by an iterative procedure. B.R. Zutter<sup>1</sup> has written an extensive FORTRAN program for the U.S.D.A. Forest Service which computes MLE's for the Weibull and this program was used in the current investigation.

---

<sup>1</sup>B.R. Zutter is a procurement forester for Buckeye Cellulose Corp., Americus, GA.

Table 5. Probability density functions (p.d.f.'s) for distributions selected as candidates for the marginal distribution of  $k$ , the rate parameter of the Chapman-Richards growth curve.

Distribution	p.d.f.*
$S_B$	$f(x) = \begin{cases} \frac{\delta}{(2\pi)^{1/2}} \frac{\lambda}{(x-\varepsilon)(\varepsilon+\lambda-x)} \exp\left\{-\frac{1}{2}\left[\gamma + \delta \log\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right)\right]^2\right\}, \\ \varepsilon < x < \varepsilon + \lambda; \delta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \varepsilon < \infty. \\ \\ 0, \text{ elsewhere.} \end{cases}$
lognormal	$f(x) = \begin{cases} \frac{(\delta) \exp\left\{-\frac{1}{2}[\lambda + \delta \log(x)]^2\right\}}{(2\pi)^{1/2}(x)}, \quad -\infty < \lambda < \infty; \delta > 0; x > 0. \\ \\ 0, \text{ elsewhere.} \end{cases}$
Weibull	$f(x) = \begin{cases} (c/b) (x/b)^{c-1} \exp[-(x/b)^c], \quad x > 0; b > 0; c > 0. \\ \\ 0, \text{ elsewhere.} \end{cases}$

\*  $\log$  = logarithm to the base  $e$ .  
 $\exp(x)$  =  $e$  raised to the power  $x$ .

The  $S_B$  distribution has four parameters. If  $k$  is the random variable, the four parameters are  $k_{min}$ ,  $k_{max}$ ,  $a$ , and  $b$ . The parameter  $k_{min}$  is the minimum value of  $k$  possible and  $k_{max}$  the maximum. The MLE's for this distribution must also be solved for iteratively. However, if  $k_{min}$  and  $k_{max}$  are known, then, conditional upon  $k_{min}$  and  $k_{max}$ , the MLE's for  $a$  and  $b$  exist in closed form (Schreuder and Hafley, 1977). In order to avoid the iterative process of estimating the four parameters of the  $S_B$  simultaneously,  $k_{min}$  and  $k_{max}$  were fixed. The estimates for  $k$  ranged from 0.000042 to 0.017. Thus 0 was deemed to be a logical value for  $k_{min}$ . Three values, 0.018, 0.02, and 0.022 were tried for  $k_{max}$ . Therefore three pairs of MLE's for  $a$  and  $b$  were found for each plot.

Next the statistic  $D$  was calculated for each of the five candidate distributions for each plot.  $D$  is the maximum absolute difference between the hypothetical cumulative distribution (e.g. the Weibull) and the observed cumulative distribution, i.e.,

$$D = \max_k |F_n(k) - F(k)|$$

where:  $F_n$  is the observed cumulative distribution

$F$  is the hypothetical cumulative distribution.

The values of  $D$  from the five distributions were then ranked for each plot, and the ranks summed over plots for the

distributions. Additionally a weighted difference  $D_w$  was calculated for each distribution as follows

$$D_w = \{ \sum n_i D_i / \sum n_i \}$$

where:  $D_i$  = value of  $D$  for plot  $i$

$n_i$  = sample size on plot  $i$ .

Based on  $D_w$  and the summed ranks of  $D$ , one distribution was then selected to characterize the marginal distribution of  $k$ .

Linear regression equations were then developed to predict the parameters of the marginal distribution of  $k$ . The dependent variables were the estimated parameter values. In this model it was assumed that one set of  $k$  and  $m$  values characterized the growth of a tree throughout its life. Thus the independent variables in the prediction models for the parameters of the distribution of  $k$  were necessarily variables measured at a reference point in time. The independent variables were height of dominants and codominants at age 20 ( $HD_{20}$ ),  $\log(HD_{20})$ ,  $(1/HD_{20})$ , number of trees per acre at age 8 ( $TS_8$ ),  $(TS_8^2)$ ,  $\log(TS_8)$ ,  $(1/TS_8)$ , and  $(TS_8 * HD_{20})$ . All logarithms were to the base  $e$ . Height of dominants at age 20 rather than the conventional base age of 25 was used because 20 was the oldest age at which all plots were measured. Trees surviving at age 8 was used because age 8 was the earliest age at which all of the thinning

treatments were completed. It was hypothesized that the differences in the distributions of  $k$  from plot to plot could be accounted for by either variable site quality, as measured by  $HD_{20}$ , by the thinning treatment or by the various transformations of these two variables.

Using the independent variables listed previously,  $R^2$  values were calculated for each possible linear prediction equation for each parameter. The models which gave the highest  $R^2$  values were then evaluated on the basis of PRESS and SUMABS. A prediction model was then chosen for each parameter.

#### 4.2.5 Prediction of $m$

For those plots on which  $m$  was not assumed to be zero, it was necessary to develop a method for predicting  $m$ . Rather than assuming the parameters of the growth curve to be independent as was done by Krause et al. (1967), the known dependence between the parameters was utilized. This was done by regressing  $m$  on  $k$  and  $k^2$ . The variables  $TS_8$  and  $HD_{20}$  were also included to allow for differences in the relationship between  $k$  and  $m$  between plots. The final model was chosen after considering  $R^2$ , PRESS, SUMABS, and SUMRES.

#### 4.2.6 Survival function

A survival curve was needed to complete the Parameter Distribution model. In this analysis, the survival curve presented by Smalley and Bailey (1974) was modified. The original Smalley and Bailey survival function was of the form

$$\log(TP/TS) = AP \{ a \log(TP) + b HD + c\sqrt{HD} \}$$

where: AP = age since planting

TP = number of trees planted per acre

TS = number of trees surviving per acre at age AP

HD = mean height of dominants and codominants at age AP.

This function has the desirable characteristic that at the age of planting AP=0, and thus TP=TS. The curve has also been used by Feduccia et al. (1979) and studied by Burkhart et al. (1981). It seems to give reasonable estimates for surviving numbers of stems (Feduccia et al., 1979; Burkhart, pers. comm.). The function was modified in the present study to the form

$$\log(TRT/TS) = (AP - AT) \{ a \log(TRT) + b HD + c\sqrt{HD} \}$$

where: TRT = the number of stems per acre left after thinning treatment

AT = age since thinning treatment

The other variables remain as defined above. In this form the curve takes into account the different thinning treatments. This form also retains the desirable characteristic that when  $AP=AT$ ,  $TS=TRT$ .

#### 4.2.7 Application

The Parameter Distribution model was applied on the fourth replication. If the model were to perform poorly, this would be most evident at the later ages encompassed by the data. Thus for each plot on replication four, the last age at which that plot was measured was selected for testing the model. Using these ages, the observed mean heights of dominants and codominants, and the actual numbers of trees left after thinning, estimated parameter values were computed for the marginal distributions of  $k$  via the prediction models developed on the fitting data set. The number of trees per acre surviving at the last age on each plot of the validation set was predicted using the fitted survival function. For comparison purposes, this number was then reduced to the exact acreage of the given plot. Using the reduced number,  $k$  values were then randomly drawn from the estimated marginal distribution of  $k$  for each stem predicted to be present. For plots 43, 44, and 45,  $m$  values were then predicted for each  $k$  value using the fitted

regression of  $m$  on  $k$ ,  $k^2$ , and  $TS_8$  (the regression chosen did not include  $HD_{20}$ ). These  $m$  values estimated the mean  $m$  for a given pair of  $k$  and  $TS_8$  values, and were called deterministic. A second pair of  $k$  and  $m$  values were generated by adding a  $N(0, \sigma^2)$  random deviate to the deterministic  $m$  value. The variance of this random deviate was taken to be equal to the MSE from the fitted regression of  $m$  on  $k$ ,  $k^2$ , and  $TS_8$ . These  $m$  values were called stochastic. All predicted negative  $m$  values, deterministic or stochastic, were set to zero. For plots 41 and 42  $m$  was set to zero for each tree predicted to be present after adjusting the estimate from the survival curve to the proper acreage. These values were called deterministic and no stochastic  $m$  values were computed. Once  $k$  and  $m$  were known (predicted) the predicted basal area was determined. Thus for each of plots 41 and 42 there was one predicted basal area distribution, while for each of plots 43, 44, and 45 there were two.

For every predicted basal area distribution, a predicted height distribution was generated. This was done by predicting a height for each predicted individual tree basal area, using the fitted height-basal area model (see section 4.1). Inputs to this model were age, observed mean height of dominants and codominants at age 20, predicted

number of stems per acre, predicted plot basal area per acre (obtained by summing the predicted individual tree basal areas and converting to per acre values), and the predicted basal area of the individual tree. One set of heights was generated by using the mean heights predicted from the regression. These heights were called deterministic. Another set was obtained by adding a  $N(0, \sigma^2)$  random variable to the deterministic height. The variance of the random variable was taken to be equal to the MSE from the fitting of the height model on the first three replications.

Every predicted distribution was compared to the appropriate observed distribution. The comparisons were done subjectively by inspecting histograms of the observed and predicted distributions, and by looking at the means and standard deviations of the predicted distributions in relation to the observed. Objective comparisons were also performed via the two-sample Kolmogorov-Smirnov (K-S) test. This is a non-parametric statistical test designed to test whether two independent samples were drawn from the same population.

As the results obtained from a stochastic simulation model vary according to the initial random number seed, the model was run on replication four with twenty different randomly selected seeds. The model was also run once on

each of the first three replications in order to determine if it behaved differently on the fitting than on the independent validation data.

### 4.3 NONLINEAR GROWTH FUNCTION

#### 4.3.1 Specification of model

Assuming the parameters of the C-R curve to be constant throughout the life of a tree, and developing a methodology to predict them, as done in the previous section, represents one means of applying this function to individual tree growth. Another method involves modeling the parameters of the curve as functions of tree- and stand-level variables. This approach has the advantage that the growth curve predicted for a stem would change as the variables which model the parameters of the curve change. At any point in time the model would predict the subsequent growth of the stem to follow a C-R growth curve, with the parameters of the curve based on the latest available information (tree- and stand-level variables measured at the current time). One form of such a model is

$$BA_{i+1} = A(1 - e^{-f_i(AP_{i+1}-c)})g_i \quad (3)$$

where:  $BA_i$  = basal area at time  $i$ .

$AP_i$  = age from planting at time  $i$

$f_i, g_i$  = functions of tree- and stand-level variables at time  $i$

$A$  = asymptote

$c$  = constant

$e$  = base of natural logarithms.

In model (3) BA tends to A as AP tends to infinity, and BA = 0 when AP = c. Since f and g are not specified, (3) actually represents a class of models.

This class of models describes the growth of an individual tree and permits the tracking of an individual tree over time, unlike the parameter distribution model which yields basal area distributions at points in time but does not permit following the growth of an individual tree.

In order to ascertain whether this approach would result in improved growth predictions over more conventional methods, a model belonging to the class of models in (3) was developed. The model was then compared to a linear model of individual tree growth on an independent data set.

The variables k and m of the C-R curve are constant throughout the life of a tree when the model is fit in the form of (1), so it was not possible to check directly the relationships of these parameters to time-dependent tree and stand-level variables which are constantly changing throughout the life of an individual tree. Furthermore, the data set was too small to check the relationships by examining the estimated parameter values for many trees at different levels of a given variable at a reference point in time while holding the other stand variables constant. Thus the functions f and g which model k and m of the C-R curve respectively were hypothesized. The functions chosen were:

$$f_i = (a CR_i + b B_i + c APA_i + d BA_i + e TS_i)^2$$

$$g_i = (a + b TS_i)$$

where:  $CR_i$  = predicted crown ratio at time  $i$

$B_i$  = plot basal area in sq. ft. at time  $i$

$APA_i$  = Area Potentially Available to the subject tree  
at time  $i$

$TS_i$  = No. of trees surviving per acre at time  $i$

$BA_i$  = Basal area in sq. ft. of the subject tree at  
time  $i$ .

$a, b, c, d, e$  are coefficients to be estimated.

$B$ ,  $TS$ , and  $BA$  are measured variables and need no explanation. Crown ratio ( $CR$ ) is defined as the length of live crown divided by total height. This variable was not measured on the data set. However, Feduccia et al. (1979) presented a model to predict crown ratio for individual loblolly pines based upon the tree's d.b.h., quadratic mean diameter for the stand, average height of dominants and codominants in the stand, and age. The data for their study came from plantations of similar origin and geographical region as those used in the present study. The mean predicted crown ratios for all plot-age combinations on the first replication are shown in Table 6. The values seem reasonable except at the lowest density, at which there

appears to be a positive bias in the predictions. Daniels (1981) used this model to predict crown ratios for the data used in the present study, and found that inclusion of predicted crown ratio improved the fit of his basal area increment model. Thus the model was accepted despite the apparent problem at the lowest density. Hahn and Leary (1979) have used predicted crown ratio in their diameter increment model. Thus there is a precedence for this procedure.

Area potentially available (APA) is an individual tree-based variable designed to quantify the effects of competition from neighboring trees on a subject tree. This variable measures the horizontal area assumed to be available to the subject tree. It is computed from a mathematical expression relating tree diameter to diameter of and distance to adjacent trees (Moore et al., 1973). Several forms of this variable have been presented in the literature. The primary difference between them is in the function which is used to weight inter-tree distances. Daniels (1981) found that of all the weighting factors presented thus far, the ratio

$$\{D_s^2 / (D_s^2 + D_c^2)\}$$

where:  $D_s$  = d.b.h. of subject tree

$D_c$  = d.b.h. of competitor

Table 6. Mean estimated crown ratios for loblolly pine trees, predicted with the equation by Feduccia et al. (1979), by plot and age for replication one of the Hill Farm data.

Age	Plot				
	11	12	13	14	15
6	77.8	84.5	---*	---	---
7	---	76.8	---	---	---
8	70.5	69.6	68.0	---	---
9	59.0	65.0	74.4	62.8	46.2
10	54.7	60.3	70.1	76.4	88.9
13	50.7	56.0	66.3	72.1	84.8
20	42.6	47.2	57.4	63.3	75.5
21	---	38.4	44.7	49.9	---

Density at age 8 (stems/acre)	1046	641	317	211	103
-------------------------------------	------	-----	-----	-----	-----

\* Plot not measured that year.

yielded the highest correlation between APA and basal area increment. Daniels (1981) also found that APA had a higher correlation with basal area increment than any other individual tree competition index considered. Thus APA was included in (3).

It is widely believed that individual tree basal area and the density of the surrounding stand are highly correlated with its subsequent growth, therefore B, BA, and TS were included in (3). Note that if APA were a perfect measure of competitive stress upon an individual stem, stand-level measures such as B and TS would be unnecessary. However, this has not been found to be true in the past. Instead, stand-level measures of competition have been found to be statistically significant in explaining basal area growth in the presence of individual tree competition indices (Daniels, 1981).

Initially, the function  $f$  in (3) was linear, since linear models are generally reasonable approximations in the absence of any specialized knowledge. However, in the process of fitting the overall model (3), using nonlinear least squares, the function  $f$  showed a tendency to become negative. As this function models  $k$  of the C-R curve and  $k$  must be positive, the expression was squared. The function  $f$  was also modeled with an intercept term. The fit was

nearly identical in terms of  $R^2$  and MSE to that where the intercept term was not included. Thus the intercept term was dropped for simplicity. Additionally, the terms TS and B were exchanged for their inverses in one run, as it was felt that the rate parameter  $k$  should have an inverse relationship with density. Again the fit was nearly identical in terms of  $R^2$  and MSE, so the form of the function shown above was adopted.

In a separate part of this investigation it was found that the parameter  $m$  of the C-R growth curve was related to initial thinning treatment (see section 4.2.3). Thus  $m$  was modeled as a simple linear regression involving number of surviving stems per acre as the independent variable.

Due to the inclusion of APA in (3), the model is distance-dependent, and thus would find application in distance-dependent individual tree models.

#### 4.3.2 Application

Model (3) was fit on the first three replications of the data set using nonlinear least squares. All the thinning treatments were fit at once, since the parameters of the model should theoretically take the different stand conditions into account. Whenever there were two subsequent measurements made on an individual stem, this constituted

one observation. The dependent variable was the basal area at the second measurement, and the independent variables were CR, TS, B, BA, and APA at the first. AP in the expression  $(AP - c)$  was taken to be age from planting at the second measurement. The asymptote A and the parameter c were fixed at 22.3 square feet and 3 years, respectively.

The model was tested on the fourth replication. Whenever there were two subsequent measurements made on a tree, the model was applied using the values of the independent variables at the first measurement to predict the basal area of the tree at the second. The differences and absolute differences between observed and predicted basal areas were then summed and averaged over replication four. For comparison, a linear model of the form

$$BA_{i+1} = b_0 + b_1 CR_i + b_2 APA_i + b_3 BA_i + b_4 B_i + b_6 TS_i + b_7 (AP_{i+1} - AP_i)$$

where the variables remain as previously defined was fit on the third replication with ordinary least squares and tested on the fourth using the same methods as for model (3). The linear model represented a conventional approach to the problem of predicting growth and served as a 'base line' for comparison.

In order to ascertain the effect of specifying the asymptote to be 22.3 square feet in (3), the entire

procedure was repeated with asymptotes of 18.96 (15% below 22.3) and 25.65 (15% above 22.3) square feet.

## Chapter 5

### RESULTS AND DISCUSSION

#### 5.1 HEIGHT-DIAMETER/BASAL AREA CURVE

##### 5.1.1 Choice of curve

The results from summing the ranks of the five candidate height-diameter/basal area curves based on  $R^2$ , PRESS, and SUMABS over the 105 plot-age combinations on the first three replications are shown in Table 7. Model 1, which predicts mean height, conditional upon basal area was clearly the best of the five on the data used in this study. It may be argued that the test was unfairly biased in favor of model 1 since the criteria were based on predicted heights, and the other models predict  $\log(H)$  or  $\log(H-4.5)$ . However, the author feels that this was completely appropriate, since height and not its logarithm is the quantity of interest. The model which predicted height directly thus had a built-in advantage over the others, and the test should have reflected this.

Curtis (1967) suggested that models which predict height rather than  $\log(H)$  should perform poorly on data sets which include many young trees. In particular, such models will likely predict negative heights for small basal areas. The data used in the current study was from young stands and

Table 7. Summed ranks over 105 plot-age combinations on the Hill Farm data for five candidate height-diameter (basal area) models, based on three statistics.\*

Model	$R^{2**}$	PRESS	SUMABS
1. $H = a + b \log(\text{BA})$	237 (1) <sup>#</sup>	226 (1)	225 (1)
2. $\log(H) = a + b \log(\text{BA})$	311 (2)	302 (2)	310.5 (2)
3. $\log(H-4.5) = a + b \log(\text{BA})$	386 (5)	380 (5)	<b>393</b> (5)
4. $\log(H) = a + b (1/D)$	315 (3)	332 (3)	314 (3)
5. $\log(H-4.5) = a + b (1/D)$	326 (4)	335 (4)	332.5 (4)

\* Statistics ranked such that 1 = best, . . . . ., 5 = worst for a particular plot-age combination.

\*\*  $R^2$ , PRESS, and SUMABS transformed to H for  $\log(H)$  and  $\log(H-4.5)$  models.

# Figure in parentheses is relative rank of summed rank.

yet model 1 was the best under all three criteria. Note especially that this model was best under PRESS and SUMABS. If the model had predicted negative heights for small basal areas, these statistics would have magnified the problem. Apparently, the danger of predicting negative heights for low basal areas when height is the dependent variable is real only when low basal areas are not well represented in the fitting data. In the Hill Farm data, average basal areas on the first three replications at ages 5 and 6 were 0.029 and 0.049, respectively, indicating that small basal areas were well represented. The heights associated with these basal areas were necessarily positive and kept the regression line in the range of positive values. Perhaps at basal areas smaller than those in the data set the model would predict negative heights; however, basal areas so small would be of little practical interest.

Two other aspects of these results are noteworthy; first, the intuitively appealing feature of modelling  $\log(H-4.5)$  rather than  $\log(H)$  did not improve the two  $\log(H)$  models' performance. Since data sets which include many young stems, such as the one used here, are where this adjustment should logically be most important, it seems that the adjustment is unnecessary. Second, the widely used  $\log(H)$ -reciprocal of diameter model (model 4) was found to

be a poorer predictor of mean height than model (1) or model (2), the allometric model. Although this study used only one data set, and this set was limited in age range, it certainly appears that the relative performance of these models should be tested further.

#### 5.1.2 Choice of surrogate regressions

The coefficients  $a$  and  $b$  of model 1 were individually regressed upon all possible linear combinations of the following variables:

1. age since planting, (AP)
2. age since thinning treatment, (AT)
3. quadratic mean diameter at age AP, (D)
4. plot basal area per acre at age AP, (B)
5. average height of dominants and codominants at age AP, (HD)
6.  $\log(\text{HD})$
7. number of trees surviving per acre at age AP, (TS)
8.  $\log(\text{TS})$
9. thinning treatment, (TRT)

All logarithms were to base  $e$ . After fitting all the possible models for each coefficient, the  $R^2$  values were examined. Those models which performed well under  $R^2$  and had the fewest independent variables were selected for

further examination with PRESS, SUMABS, and SUMRES. For coefficient a, seven models were chosen for further investigation. These seven models are shown in Table 8. Model 7a was the best in terms of R<sup>2</sup>, PRESS, SUMABS, and third under SUMRES. The choice was clear, and model 7a was selected as the surrogate for a. Six models were chosen as candidates for the surrogate for b. They are listed, along with their R<sup>2</sup>, PRESS, SUMABS, and SUMRES values in Table 8. Here the choice was less apparent. Models 3b and 6b both performed well, with 3b being the best under PRESS and SUMRES and fourth under R<sup>2</sup> and SUMABS. Model 6b was best under R<sup>2</sup> and SUMABS, second under PRESS and fifth under SUMRES. Between the two models the only statistic that was very different from a practical standpoint was SUMRES. SUMRES is the sum of the PRESS residuals, and gives an indication of a model's bias. Model 3b had a SUMRES of -0.056, suggesting very little bias, while model 6b had a SUMRES of -1.496, about twenty-five times greater. On the basis of this and the fact that model 3b has one less variable than 6b and is therefore simpler, model 3b was selected.

The overall height model was thus of the following form:

$$H_i = b_0 + b_1 \log(HD_i) + b_2 D_i + b_3 HD_i + \log(BA_i) \{ b_4 + b_5 AT_i + b_6 \log(HD_i) + b_7 HD_i \}$$

Table 8. Statistics from fitting the candidate surrogate regressions for a and b in the model  $H = a + b \log(\text{BA})$  on the first three replications of the Hill Farm data.

Model	Dependent variable	Independent variables*	$R^2$	PRESS	SUMABS	SUMRES
1	a	HD	0.9428	1440.99	304.71	-0.758
2	a	HD, B	0.9642	917.20	232.60	-1.362
3	a	HD, D	0.9747	640.42	200.38	-1.532
4	a	HD, BA	0.9763	599.26	199.62	0.260
5	a	HD, BA, log(TS)	0.9771	588.03	186.25	-0.276
6	a	HD, D, AT	0.9778	578.68	190.59	-0.547
7	a	HD, D, log(HD)	0.9789	537.20	177.71	-0.312
8	b	HD, AT, log(TS)	0.6622	833.53	198.40	-0.819
9	b	HD, B, log(TS)	0.6662	814.85	190.95	0.199
10	b	HD, AT, log(HD)	0.6736	782.77	193.62	-0.056
11	b	HD, AT, log(HD), D	0.6862	845.86	202.76	-3.475
12	b	HD, AT, log(TS), B	0.6898	795.32	191.38	-0.517
13	b	HD, AT, log(TS), log(HD)	0.6916	784.81	189.66	-1.496

where the subscript  $i$  refers to age since treatment, and the variables are as described above. This model was fit to all the height-basal area pairs on the first three replications and the resulting equation was

$$H_i = -19.051 + 15.006 \log(HD_i) - 28.530 D_i + 0.752 HD_i + \\ \log(BA_i) \{-3.838 - 0.109 AT_i + 2.015 \log(HD_i) + 0.153 HD_i\} \\ R^2 = 0.97 \quad Sy.x = 3.370 \quad \text{mean } H = 30.5 \quad n = 5912$$

All the estimated coefficients were significant at the .01 probability level. The  $R^2$  value indicates that the model fit the data very well.

Finally, note that in the model

$$H = a + b \log(BA)$$

the coefficient  $a$  represents the estimated mean height for a tree with a basal area of one square foot (approx. 13.5 inches d.b.h.). However, since few trees on the data set were this large, no attempt was made to interpret  $a$  in this manner.

## 5.2 PARAMETER DISTRIBUTION MODEL

### 5.2.1 Fitting of growth curves

The average  $R^2$  values from the fitting of the C-R curve to all the stems on the first three replications are shown by plot in Table 9. For plots 11, 12, 21, 22, 31, and 32 the curve

$$\log\{1 - (BA/22.3)\} = -k(t-c)$$

was fit as a no intercept linear model and the average  $R^2$  values shown in Table 9 are in terms of  $\log\{1-(BA/22.3)\}$ . The linear form was used to simplify the estimation process. For plots 13, 14, 15, 23, 24, 25, 33, 34, and 35 the curve

$$BA = 22.3(1 - e^{-k(t-3)})^{[1/(1-m)]}$$

was used. Since this is a nonlinear model,  $R^2$  per se cannot be calculated. However, one may easily compute

$$R_{NL}^2 = \{RSS - (TSS - CTSS)\} / CTSS$$

where:  $RSS = \sum Y_i^2 - \sum (Y_i - \hat{Y}_i)^2$

$$TSS = \sum Y_i^2$$

$$CTSS = \sum (Y_i - \bar{Y})^2.$$

The statistic  $R_{NL}^2$  expresses the percent of variation explained by the regression, over and above that explained by the mean of the dependent variable.

The results in Table 9 show that the curve(s) fit the data well. Since the  $R^2$  values for the two fitting methods were computed differently, they are not directly comparable. Within each fitting method there appears to be a trend, with mean  $R^2$  values increasing as residual density following treatment decreases. This is reasonable. One might expect an individual stem to trace a slightly smoother growth curve as residual density decreases since one variable (competition) which affects growth is reduced in magnitude.

Table 9. Mean and standard deviation of  $R^2$  values from fitting the Chapman-Richards growth curve to all loblolly pines on a plot in the Hill Farm data, by plot.

Plot	mean $R^2$	standard deviation	n
11	0.77	0.4745	248
12	0.84	0.4273	150
21	0.78	0.3310	249
22	0.85	0.1925	150
31	0.82	0.3135	246
32	0.87	0.1824	150
	<u>mean <math>R_{NL}^2</math></u>		
13	0.94	0.1790	75
14	0.96	0.0707	50
15	0.99	0.0109	24
23	0.96	0.0730	74
24	0.90	0.2568	52
25	0.99	0.0094	24
33	0.98	0.0333	74
34	0.97	0.1186	50
35	0.99	0.0163	24

\* For plots 11, 12, 21, 22, 31, and 32  $R^2$  values are in terms of  $\log(1 - BA/22.3)$ , from no-intercept linear model. For plots 13, 14, 15, 23, 24, 25, 33, 34, and 35  $R_{NL}^2$  values are from the nonlinear model and are computed as follows:

$$R_{NL}^2 = \frac{\sum(Y_i - \bar{Y})^2 - \sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Thus a growth curve should tend to fit stems in stands of low density better.

### 5.2.2 Estimating the marginal distribution of k

Next the estimates for k were fit to the lognormal, Weibull, and  $S_B$  ( $k_{max} = 0.018, 0.02, \text{ and } 0.022$ ) distributions. The value of the statistic  $D = \max_k |(F_n - F)|$  which resulted from fitting each distribution to the estimated coefficients on each plot is shown in Table 10. The D values were ranked from lowest to highest for each plot and the frequency of ranks for each distribution is shown in Table 11. The values in Table 10 indicate that the Weibull, lognormal, and  $S_B(k_{max}=0.018)$  are the top three candidates. As a further test, average D values weighted by sample size were calculated for each distribution. The weighted average D values were:

lognormal	.0908
Weibull	.0663
$S_B, k_{max}=.018$	.0904
$S_B, k_{max}=.02$	.0905
$S_B, k_{max}=.022$	.0909.

Since the Weibull distribution had the lowest weighted average D, it was selected. As a further check on the validity of the Weibull, one-sample K-S tests were performed

for each plot. This test is designed to test whether a sample could have been drawn from a specified hypothetical distribution. The test is based on the statistic  $D$ . The  $p$ -levels which result from the test estimate the probability of observing a  $D$  value as large or larger than that actually observed, given that the sample did arise from the hypothetical distribution. Note however that these probabilities cannot be strictly interpreted in this case. One of the fundamental assumptions of the K-S test is that the observations in the sample are independent. Since the estimates for  $k$  come from fitting a growth curve to trees on the same plot, and thus probably influencing one another's growth, the estimated  $k$ 's are probably not independent. However, the  $p$ -levels may be regarded as an index of the 'goodness' or 'closeness' of fit. A large  $p$ -level represents a good fit, whereas a small  $p$ -level reflects the opposite. The  $p$ -levels from performing the one-sample K-S test on the fit of the Weibull distribution to the marginals of  $k$  are shown by plot in Table 12. As these values are generally high, with only one being below .05, the conventional statistical significance level, it was concluded that the Weibull estimated the marginal distributions of  $k$  adequately.

Table 10. Values and ranks for  $D = \max(D^+, |D^-|)$  from fitting the five candidate distributions for the marginal distribution of  $k$ , the rate parameter of the Chapman-Richards growth curve, to estimates of  $k$  resulting from fitting the curve to individual loblolly pine trees on three replications in the Hill Farm data.\*

Plot	n	lognormal	Weibull	$S_B$		
				kmax=0.022	kmax=0.02	kmax=0.018
11	248	0.1184 (5)	0.0450 (1)	0.1161 (4)	0.1159 (3)	0.1156 (2)
12	150	0.1000 (5)	0.0493 (1)	0.0978 (4)	0.0975 (3)	0.0972 (2)
13	75	0.1218 (1)	0.1644 (5)	0.1291 (2)	0.1299 (3)	0.1309 (4)
14	50	0.0668 (1)	0.1057 (5)	0.0737 (2)	0.0747 (3)	0.0758 (4)
15	24	0.1423 (1)	0.1900 (2)	0.1930 (3)	0.2043 (4)	0.2318 (5)
21	249	0.0738 (5)	0.0601 (1)	0.0730 (4)	0.0729 (3)	0.0728 (2)
22	150	0.0865 (5)	0.0430 (1)	0.0836 (4)	0.0832 (3)	0.0828 (2)
23	74	0.0965 (1)	0.1165 (5)	0.1003 (2)	0.1005 (3)	0.1008 (4)
24	52	0.1123 (1)	0.1572 (5)	0.1334 (2)	0.1360 (3)	0.1396 (4)
25	24	0.1467 (5)	0.0925 (1)	0.1190 (4)	0.1159 (3)	0.1121 (2)
31	246	0.0588 (5)	0.0387 (1)	0.0571 (4)	0.0569 (3)	0.0567 (2)
32	150	0.0993 (5)	0.0428 (1)	0.0981 (4)	0.0979 (3)	0.0977 (2)
33	74	0.0479 (4)	0.0823 (5)	0.0419 (3)	0.0410 (2)	0.0405 (1)
34	50	0.1150 (5)	0.0641 (1)	0.0985 (4)	0.0965 (3)	0.0939 (2)
35	24	0.0901 (2)	0.0979 (4)	0.0891 (1)	0.0924 (3)	0.0987 (5)

\* Figures in parentheses are ranks of  $D$  values within plots.

Table 11. Frequency of ranks within plots for  $D = \max(D^+, |D^-|)$  from fitting the five candidate distributions for the marginal distribution of  $k$ , the rate parameter of the Chapman-Richards growth curve, to estimates of  $k$  resulting from fitting the curve to individual loblolly pine trees on three replications in the Hill Farm data.

Rank	$S_B$				
	lognormal	Weibull	$k_{\max} = 0.022$	$k_{\max} = 0.020$	$k_{\max} = 0.018$
1	5	8	1	0	1
2	1	1	4	1	8
3	0	0	2	13	0
4	1	1	8	1	4
5	8	5	0	0	2

Table 12. Probabilities from one-sample K-S test of the fit of the Weibull distribution to estimates of  $k$ , the rate parameter of the Chapman-Richards growth curve, by plot for three replications in the Hill Farm data.\*

Plot	D**	n	p-level <sup>#</sup>
11	0.0450	248	0.6960
12	0.0493	150	0.8599
13	0.1644	75	0.0347
14	0.1057	50	0.6315
15	0.1900	24	0.3516
21	0.0601	249	0.3293
22	0.0430	150	0.9442
23	0.1165	74	0.2680
24	0.1572	52	0.1529
25	0.0975	24	0.9864
31	0.0387	246	0.8538
32	0.0428	150	0.9465
33	0.0823	74	0.6974
34	0.0641	50	0.9864
35	0.0979	25	0.9755

\* Reject the hypothesized distribution if p-level falls below a specified significance level (e.g., 0.05 or 0.01).

\*\*  $D = \max(D^+, |D^-|)$ .

# p-levels are approximate for  $n$  less than 80.

Next prediction equations for the two parameters  $b$  and  $c$  of the Weibull distribution were developed. The variables considered for inclusion as independent variables were number of trees per acre surviving at age 8 ( $TS_8$ ) and average height of dominants and codominants ( $HD_{20}$ ), and various transformations of these two quantities. These variables were entered as independent variables in a linear model, with the dependent variable being  $b$  or  $c$ . All possible combinations of the independent variables were examined for each dependent variable. For each parameter, the models with the fewest independent variables which performed well under  $R^2$  were selected for further examination with PRESS, SUMABS, and SUMRES. The candidate models selected for each parameter are shown in Table 13. The model selected as the predictor for  $c$  was 7. This model was tied for the best on the basis of  $R^2$ , best on PRESS and SUMABS, and fourth under SUMRES. Since the primary objective of the model was prediction, PRESS and SUMABS were deemed to be the most important of the four statistics, and the choice was clear. Model 13 was chosen as the predictor for  $b$  for the same reasons. This model was tied for best under  $R^2$ , best under PRESS and SUMABS, and third under SUMRES.

Table 13. Statistics for the candidate models for prediction of the two parameters, b and c, of the Weibull distribution used to characterize the marginal distribution of k, the rate parameter of the Chapman-Richards growth curve. Models fit on, and k values estimated from three replications in the Hill Farm data.

Model	Dependent variable	Independent variables	$R^2$	PRESS	SUMABS	SUMRES
1	c	$TS_8^2$ , $TS_8^{-1}$ , $\log(TS_8)$	0.59	4.484	6.846	-0.045
2	c	$TS_8$ , $TS_8^2$ , $\log(TS_8)$	0.60	4.383	6.828	-0.057
3	c	$TS_8$ , $TS_8^2$ , $TS_8^{-1}$	0.60	4.413	6.838	-0.068
4	c	$TS_8^{-1}$ , $\log(TS_8)$ , $TS_8 \times HD_{20}$	0.67	4.098	6.696	0.172
5	c	$TS_8^{-1}$ , $HD_{20}^{-1}$ , $TS_8 \times HD_{20}$ , $\log(TS_8)$	0.74	4.065	6.658	0.126
6	c	$TS_8^{-1}$ , $TS_8 \times HD_{20}$ , $\log(TS_8)$ , $\log(HD_{20})$	0.74	4.020	6.626	0.118
7	c	$HD_{20}$ , $TS_8^{-1}$ , $TS_8 \times HD_{20}$ , $\log(TS_8)$	0.74	3.987	6.604	0.110
8	c	$TS_8$ , $TS_8^2$ , $TS_8 \times HD_{20}$ , $\log(TS_8)$	0.74	4.307	6.689	1.004
9	b	$TS_8^{-1}$ , $HD_{20}^{-1}$	0.98	$3.03 \times 10^{-6}$	$5.81 \times 10^{-3}$	$7.17 \times 10^{-4}$
10	b	$TS_8^2$ , $TS_8^{-1}$	0.99	$1.60 \times 10^{-6}$	$3.78 \times 10^{-3}$	$1.52 \times 10^{-5}$
11	b	$TS_8$ , $TS_8^{-1}$	0.99	$1.46 \times 10^{-6}$	$3.47 \times 10^{-3}$	$-4.70 \times 10^{-6}$
12	b	$TS_8^{-1}$ , $TS_8 \times HD_{20}$	0.99	$1.42 \times 10^{-6}$	$3.41 \times 10^{-3}$	$-6.22 \times 10^{-6}$
13	b	$TS_8^{-1}$ , $\log(TS_8)$	0.99	$1.38 \times 10^{-6}$	$3.27 \times 10^{-3}$	$1.96 \times 10^{-5}$

Thus the final fitted prediction equations for b and c were:

$$c = -63.4345 + 0.0612 (\text{HD}_{20}) + 1506.1385 (1/\text{TS}_8) \\ - 0.0002 (\text{TS}_8)(\text{HD}_{20}) + 10.6942 \log(\text{TS}_8)$$

$$R^2 = 0.74 \quad \text{Sy.x} = 0.3750 \quad n = 15 \quad \text{mean } c = 2.351$$

$$b = -0.0075 + 0.0010 \log(\text{TS}_8) + 1.1493 (1/\text{TS}_8)$$

$$R^2 = 0.99 \quad \text{Sy.x} = 2.646 \times 10^{-7} \quad n = 15 \quad \text{mean } b = 0.003$$

It is interesting that these parameters were so well related (judging by the  $R^2$  values) to the stand variables  $\text{HD}_{20}$  and  $\text{TS}_8$ . In a different context, when diameter distributions themselves are fit directly by the Weibull, the parameters of the p.d.f. are generally poorly related to stand variables, with  $R^2$  values for c being around 0.1 (Smalley and Bailey, 1974; Feduccia et al., 1979).

### 5.2.3 Prediction of m from k

First a simple linear model regressing m on k was tried. This yielded an  $R^2$  of .761. Addition of  $k^2$  raised the  $R^2$  value to .815. Subsequent additions of  $\text{TS}_8$  and  $\text{HD}_{20}$  (in the presence of  $\text{TS}_8$ ) increased the  $R^2$  to .825 and .828, respectively. The latter three models were then evaluated on the basis of PRESS, SUMABS, and SUMRES. They were very similar for all criteria. The three-variable model was chosen as a compromise. The fitted regression equation was:

$$m = -0.1418 + 81.4591 k - 2922.1617 k^2 + 0.0002 \text{TS}_8$$

$$R^2 = 0.825 \quad Sy.x = 0.055 \quad n = 447 \quad \text{mean } m = 0.1205$$

#### 5.2.4 Survival curve

The modified survival curve detailed in METHODS was fit to the 105 plot-age combinations on the three fitting replications with the following results:

$$\log(\text{TRT}/\text{TS}) = (\text{AP} - \text{AT}) (0.0116 \log(\text{TRT}) + 0.0023 \text{HD} - 0.0254\sqrt{\text{HD}})$$

$$R^2 = 0.74 \quad Sy.x = 0.0592 \quad n=105$$

The  $R^2$  value was inflated slightly because the regression was forced through the origin. The model was also fit without this restriction, and an  $R^2$  value of 0.70 was obtained.

These results compare favorably with those of Smalley and Bailey (1974) and Feduccia et al. (1979) who, using the curve in its original form, reported  $R^2$  and  $Sy.x$  values of 0.84 and 0.0648, and 0.66 and 0.1342, respectively.

#### 5.2.5 Application to independent data

The Parameter Distribution model was applied to the fourth replication. The results are summarized in Tables 14-17 and Figures 2-11. Tables 14 and 15 display the observed and predicted means and standard deviations of basal area and height, respectively. The values in these

tables were obtained with one run of the model, i.e., one seed, and are presented for the fitting and test data sets. These Tables also display the p-levels resulting from the comparison of the predicted and observed distributions via the two-sample K-S test. The null hypothesis associated with this test is that the two samples arose from the same population. The p-levels approximate the probability of observing a difference between the c.d.f.'s of the two samples as large or larger than that actually observed, given that the null hypothesis is true. A high p-level is thus desirable. Note however that due to the problem of independence mentioned earlier, these p-levels are only general indicators of probabilities. The p-levels in general appear to be low. However, it must be kept in mind exactly what the K-S test is testing, which is whether or not the two samples came from the same population. Because of all the abstraction involved in the parameter distribution model this may be too harsh a test. Given any set of initial stand characteristics, the final, observed distribution may be considered a random variable. The predicted distributions resulting from the parameter distribution model are also random variables. Thus one might expect some differences between the observed and predicted distributions, which would be reflected in low p-

levels. Furthermore, the K-S test is sensitive to sample size. When the samples are large (e.g.  $> 100$ ), it does not take too great a difference between the two samples to produce a low p-level. Thus a more reasonable method for judging the performance of the model is a subjective comparison of the observed and predicted means and standard deviations, and of the plotted observed and predicted distributions.

The p-levels do yield one useful piece of information. Within plots for which  $m$  was predicted through regression analysis, the predicted basal area tended to agree better with the observed when  $m$  was stochastic. A similar trend is evident in Table 15 for the height distributions. At least one of prediction methods II, III, and IV (see Table 15 for definition of method types) was better than method I in 15 of the 20 cases. However, there was no discernable trend among II, III, and IV as to which was better.

The results in Tables 14 and 15 demonstrate that the results obtained for the fitting and independent data sets were similar.

Figures 2-11 show the observed and predicted distributions resulting from one run (one seed) of the model on the independent data. The seed used in these Figures was the same as that used to compute the values in Tables 14 and

Table 14. Means and standard deviations of basal area (sq. ft.) observed on one replication of the Hill Farm data and of predicted basal area, and p-levels from comparison of observed and predicted basal area distributions via the two-sample K-S test.\*

Plot	Age	Obs.		Pred.		Means		Standard Deviations		p-level	
		n	n	n	n	I**	II**	Obs.	I	II	I
11	20	175	187	0.2443	0.2403	----	0.0985	0.1318	----	0.0240	----
12	21	109	116	0.3371	0.3280	----	0.1296	0.1180	----	0.6255	----
13	21	72	66	0.5237	0.5153	0.5255	0.1803	0.1473	0.1924	0.6379	0.3638
14	21	49	46	0.6497	0.6577	0.6715	0.1928	0.2087	0.2478	0.4202	0.5293
15	20	22	28	0.9340	0.9147	0.9806	0.1327	0.1139	0.3073	0.0987	0.3672
21	21	128	172	0.2767	0.2557	----	0.1372	0.1463	----	0.1029	----
22	21	120	105	0.3913	0.3135	----	0.1716	0.1505	----	0.0064	----
23	21	73	68	0.4543	0.5090	0.4989	0.1372	0.1563	0.2091	0.0059	0.1004
24	21	48	46	0.6154	0.6578	0.6590	0.2238	0.1892	0.2503	0.0491	0.3234
25	21	24	24	0.9690	1.0051	1.0254	0.2547	0.0739	0.2562	0.0325	0.4536
31	21	176	177	0.2290	0.2547	----	0.1074	0.1286	----	0.0152	----
32	20	143	126	0.2684	0.3124	----	0.0938	0.1022	----	0.0037	----
33	21	73	67	0.4854	0.5252	0.5391	0.1352	0.1478	0.1979	0.0117	0.0193
34	21	50	48	0.6423	0.6708	0.6940	0.1725	0.2054	0.2648	0.0258	0.1683
35	21	25	25	1.1884	1.0503	1.0717	0.3214	0.1389	0.2668	0.0066	0.7129
41	21	150	166	0.2845	0.2627	----	0.1347	0.1734	----	0.0097	----
42	21	138	120	0.2775	0.3252	----	0.1135	0.1132	----	0.0002	----
43	20	68	67	0.4872	0.4592	0.4560	0.1328	0.1562	0.1853	0.3354	0.1082
44	20	50	48	0.5666	0.6435	0.6477	0.1638	0.2064	0.2503	0.0004	0.0095
45	21	25	26	0.9912	0.9928	1.0496	0.2420	0.0600	0.1986	0.0911	0.2688

\* Predicted basal area distributions from one run of parameter distribution model with a representative seed.

\*\* Obs.: Observed.

I: Predicted, m deterministic.

II: Predicted, m stochastic.

Table 15. Means and standard deviations of height (feet) observed on one replication in the Hill Farm data and of predicted height, and p-levels from comparison of observed and predicted height distributions via the two-sample K-S test.\*

Plot	Age	Pred. n#	Means				Standard Deviations				p-level							
			Obs.**	I**	II**	III**	IV**	Obs.	I	II	III	IV	I	II	III	IV		
11	20	51	187	54.23	51.88	51.88	---	---	---	6.52	7.48	8.34	---	---	0.0360	0.0033	---	---
12	21	91	116	57.81	57.56	57.51	---	---	---	7.70	5.10	6.31	---	---	0.1296	0.3913	---	---
13	21	70	66	61.98	61.16	60.84	60.84	60.48	60.48	6.50	4.66	5.93	5.74	6.30	0.0173	0.1091	0.1899	0.0264
14	21	47	46	62.22	62.76	61.98	62.52	61.77	10.92	9.33	9.99	9.77	10.66	6.85	0.0415	0.4320	0.1922	0.6202
15	20	22	28	56.52	55.53	56.07	54.95	54.57	3.06	1.76	4.49	3.64	5.87	4.29	0.0168	0.2255	0.1768	0.0699
21	21	68	172	62.32	56.39	56.42	---	---	6.43	8.53	9.43	---	---	---	0.0001	0.0001	---	---
22	21	64	105	68.16	62.89	62.71	---	---	6.90	8.28	8.55	---	---	---	0.0003	0.0000	---	---
23	21	71	68	56.26	57.24	57.06	56.65	55.82	8.01	4.50	5.52	6.04	7.05	6.85	0.0576	0.5832	0.4963	0.6694
24	21	45	46	61.18	63.71	63.70	63.30	63.97	10.30	4.67	5.54	5.82	6.85	4.29	0.0322	0.2494	0.5390	0.5390
25	21	24	24	63.17	64.43	64.56	63.98	63.74	6.81	1.02	3.10	3.78	4.29	4.29	0.0325	0.2677	0.6888	0.6888
31	21	76	177	58.58	54.74	54.79	---	---	6.78	7.08	7.58	---	---	---	0.0101	0.0054	---	---
32	20	60	126	52.30	52.04	52.11	---	---	5.38	4.08	5.28	---	---	---	0.2224	0.1210	---	---
33	21	69	67	59.72	59.21	58.86	58.87	58.68	5.71	4.45	5.98	5.60	7.05	7.05	0.0462	0.2991	0.2905	0.1433
34	21	48	48	58.97	57.97	57.22	57.69	57.18	7.48	8.72	9.40	9.14	9.99	9.99	0.0610	0.2545	0.6969	0.1650
35	21	25	25	65.99	64.73	65.07	64.39	63.82	6.21	1.56	4.72	3.37	5.09	5.09	0.0008	0.1603	0.0383	0.0166
41	21	62	166	65.28	57.81	57.84	---	---	5.19	10.36	10.61	---	---	---	0.0000	0.0000	---	---
42	21	60	120	57.48	55.37	55.39	---	---	4.94	4.82	6.03	---	---	---	0.2091	0.1308	---	---
43	20	68	67	58.44	59.91	59.62	59.60	58.91	5.57	5.84	6.67	6.46	7.50	7.50	0.0273	0.0749	0.2327	0.4860
44	20	44	48	59.88	58.62	59.41	58.32	57.14	6.32	5.97	6.35	6.68	7.50	7.50	0.0260	0.4977	0.3541	0.2782
45	21	24	26	64.82	63.72	64.71	63.44	64.02	5.32	0.86	3.17	2.52	4.10	4.10	0.0063	0.5029	0.1709	0.3134

# - Observed n is not the number of trees surviving per plot.

\* - Predicted height distributions from one run of parameter distribution model with a representative seed.

\*\* - Obs. : Observed.

I : Predicted, m deterministic, height deterministic.

II: Predicted, m deterministic, height stochastic.

III: Predicted, m stochastic, height deterministic.

IV: Predicted, m stochastic, height stochastic.

Table 16. Mean and standard deviation of basal area (sq. ft.) observed on one replication of the Hill Farm data, and of means and standard deviations of predicted basal area based on 20 runs of the parameter distribution model.\*

Plot	Observed mean	Predicted mean (std. dev.)		Observed standard deviation	Predicted std. deviation mean (std. dev.)	
		I	II		I	II
41	0.2485	0.2523 (0.0114)	—	0.1347	0.1584 (0.0104)	—
42	0.2775	0.3261 (0.0122)	—	0.1135	0.1114 (0.0058)	—
43	0.4872	0.4797 (0.0148)	0.4744 (0.0185)	0.1328	0.1472 (0.0135)	0.1850 (0.0154)
44	0.5666	0.6375 (0.0234)	0.6371 (0.0256)	0.1638	0.1904 (0.0161)	0.2547 (0.0250)
45	0.9912	0.9932 (0.0263)	1.0039 (0.0594)	0.2420	0.1079 (0.0324)	0.2839 (0.0433)

\*Prediction systems: I - m deterministic  
 II - m stochastic

Table 17. Mean and standard deviation of height (feet) observed on one replication in the Hill Farm data, and of means and standard deviations of predicted height based on 20 runs of the parameter distribution model.\*

Plot	Observed mean	Predicted mean				Observed standard deviation	Predicted standard deviation			
		I	II	III	IV		I	II	III	IV
41	65.28	57.89 (0.51)	57.88 (0.53)	-----	-----	5.19	10.07 (0.61)	10.54 (0.66)	-----	-----
42	57.48	55.41 (0.22)	55.36 (0.30)	-----	-----	4.94	4.71 (0.58)	5.82 (0.55)	-----	-----
43	58.44	60.22 (0.30)	60.17 (0.43)	59.83 (0.34)	59.80 (0.60)	5.57	5.17 (0.92)	6.07 (0.93)	6.10 (0.80)	6.87 (0.94)
44	59.88	58.67 (0.43)	58.62 (0.57)	58.22 (0.49)	58.16 (0.79)	6.32	5.98 (1.71)	6.75 (1.63)	6.96 (1.60)	8.08 (1.47)
45	64.82	63.64 (0.06)	63.74 (0.78)	63.19 (0.16)	63.32 (0.66)	5.32	1.58 (0.62)	3.83 (0.53)	3.78 (0.61)	4.99 (0.61)

\*Prediction systems: I - m deterministic, height deterministic  
 II - m deterministic, height stochastic  
 III - m stochastic, height deterministic  
 IV - m stochastic, height stochastic

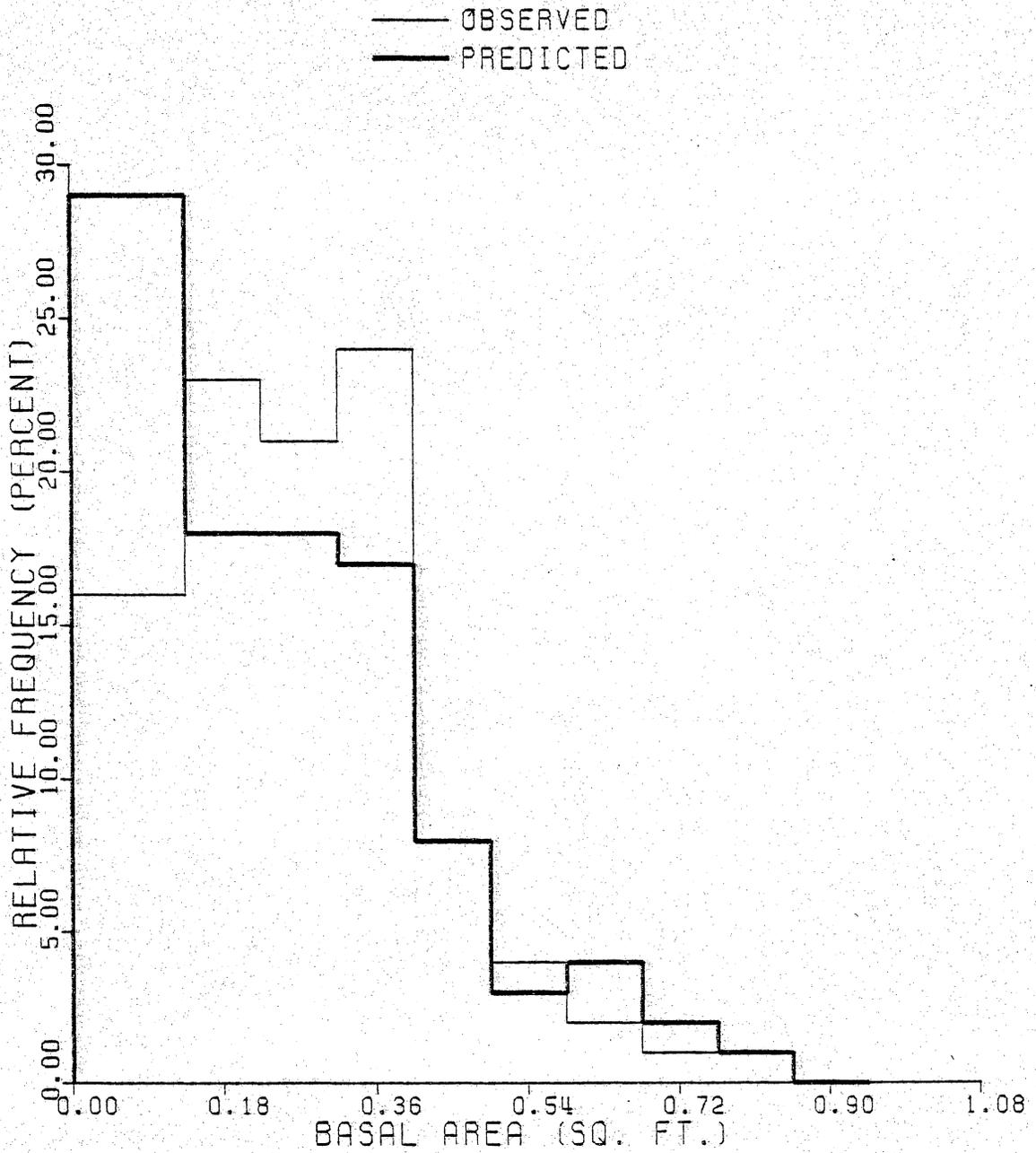


Figure 2. Observed basal area distribution at age 21 on plot 41 of the Hill Farm data, and basal area distribution predicted with the parameter distribution model.

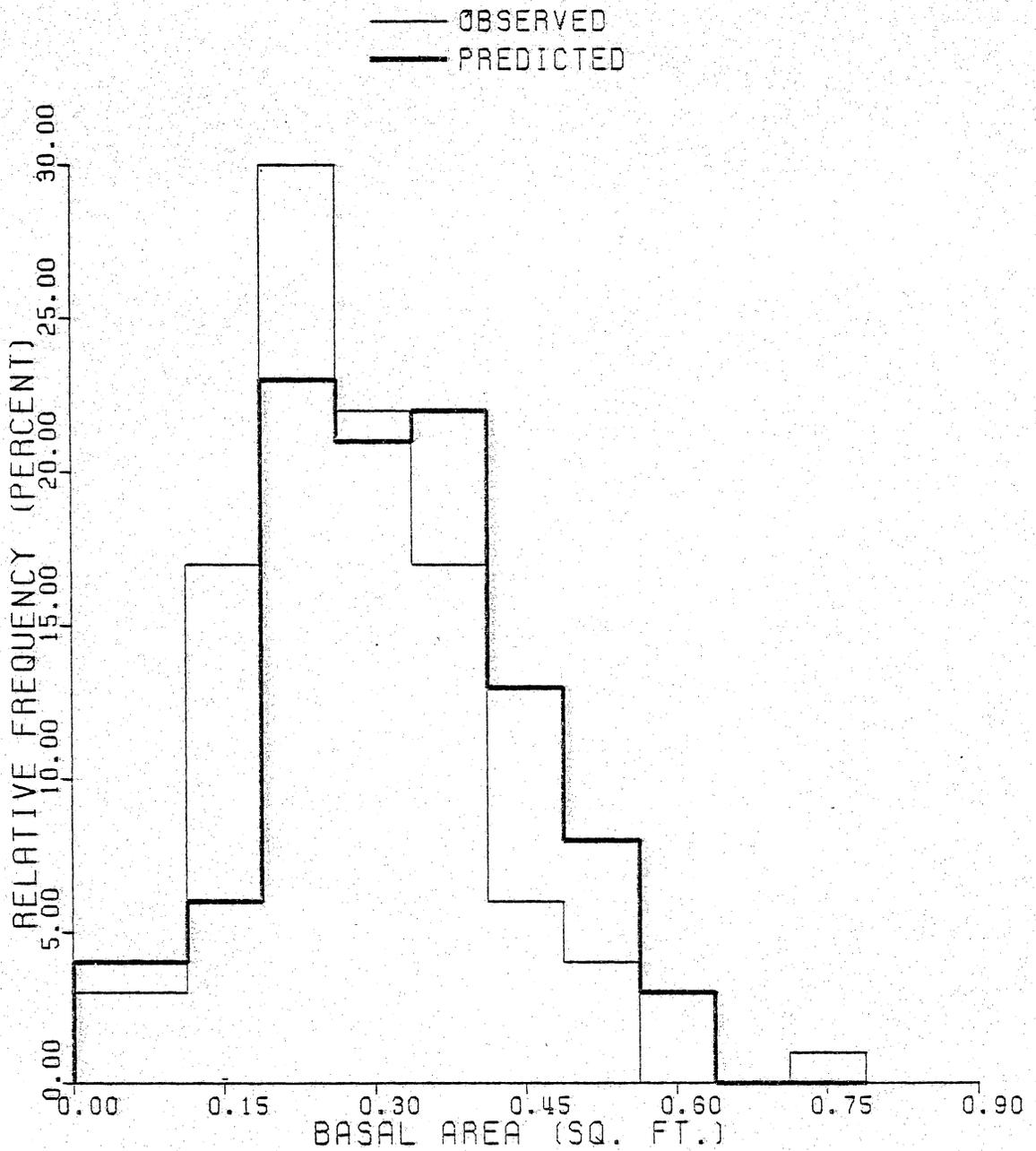
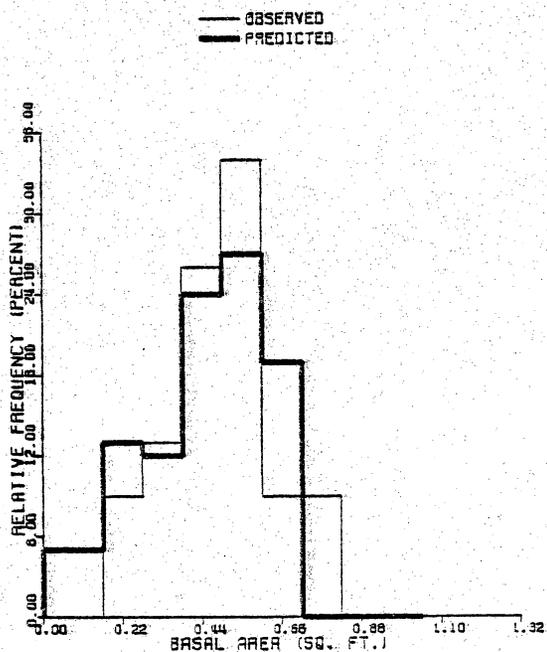
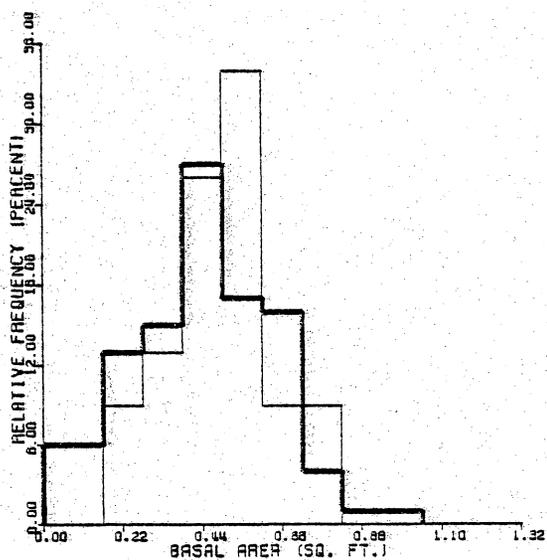


Figure 3. Observed basal area distribution at age 21 on plot 42 of the Hill Farm data, and basal area distribution predicted with the parameter distribution model.

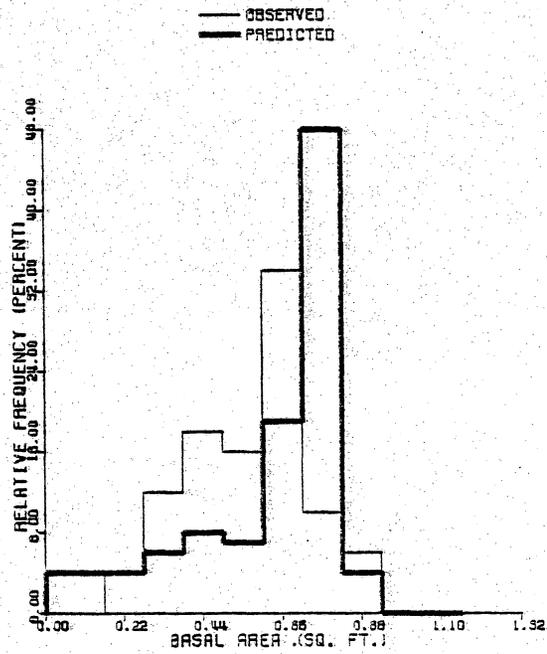


a. m deterministic

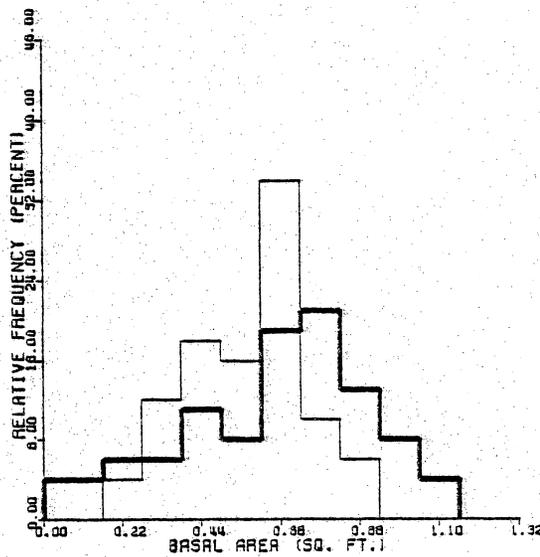


b. m stochastic

Figure 4. Observed basal area distribution at age 20 on plot 43 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model.

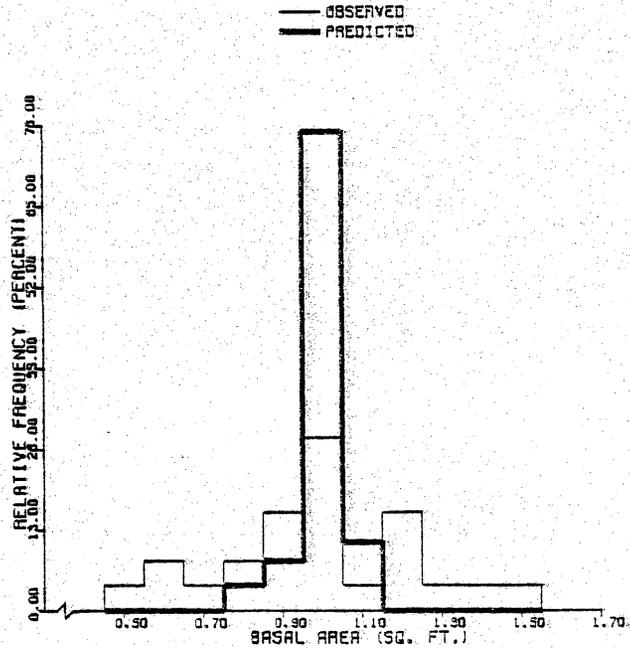


a. m deterministic

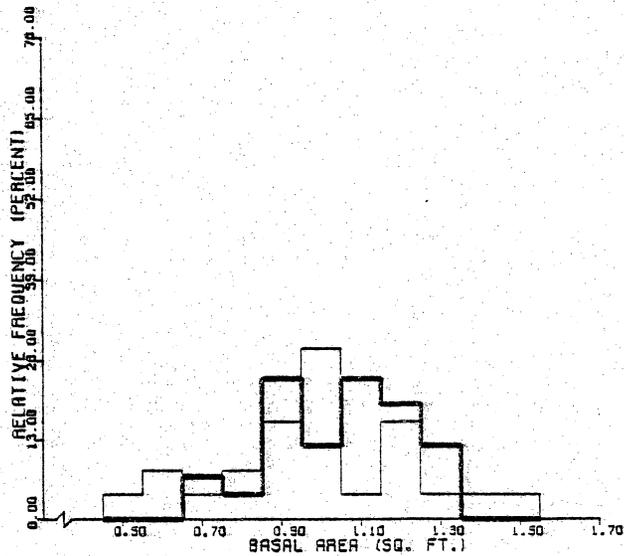


b. m stochastic

Figure 5. Observed basal area distribution at age 20 on plot 44 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model.

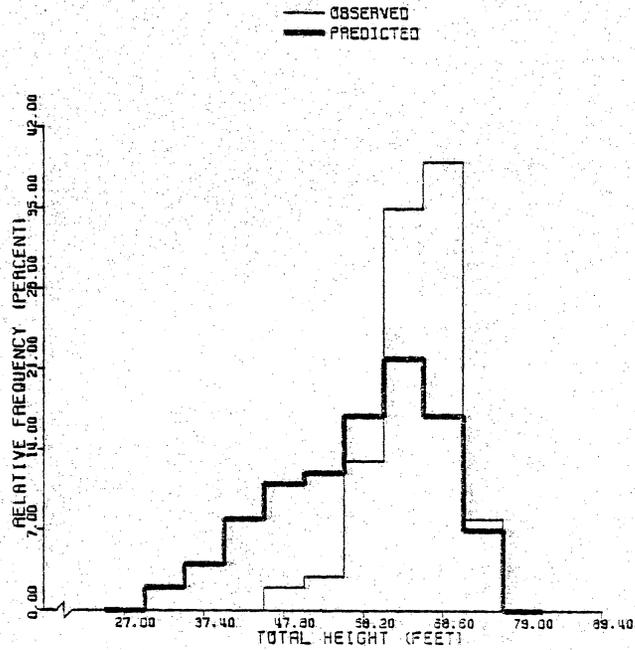


a. m deterministic

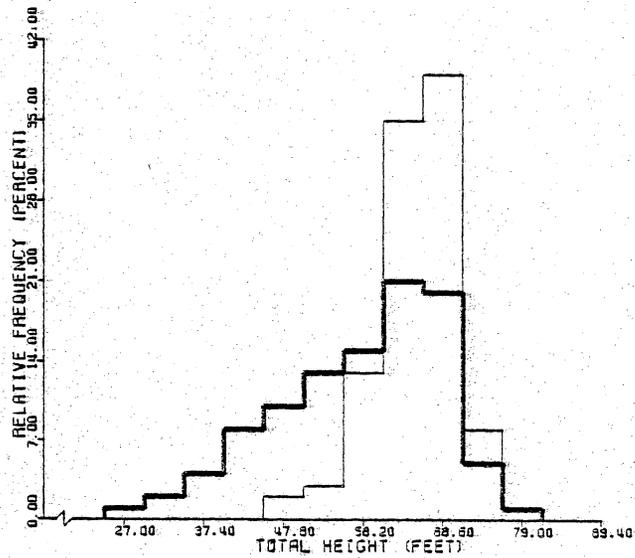


b. m stochastic

Figure 6. Observed basal area distribution at age 21 on plot 45 of the Hill Farm data, and basal area distributions predicted with the parameter distribution model.

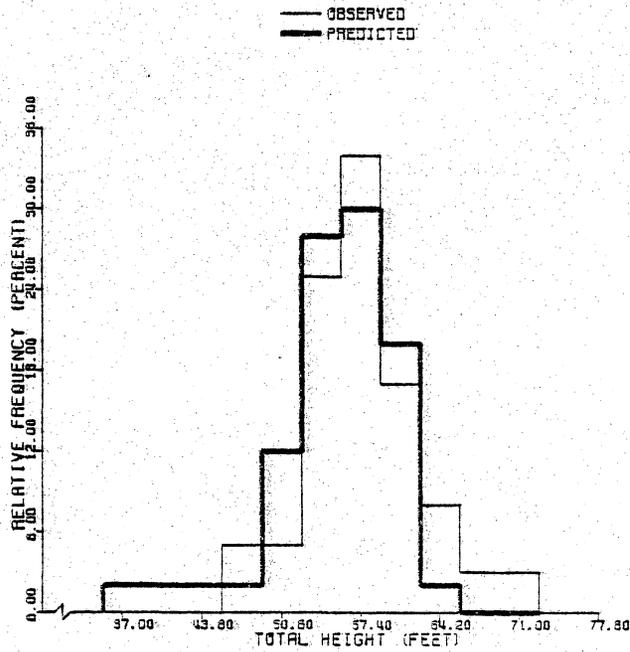


a. m deterministic  
height deterministic

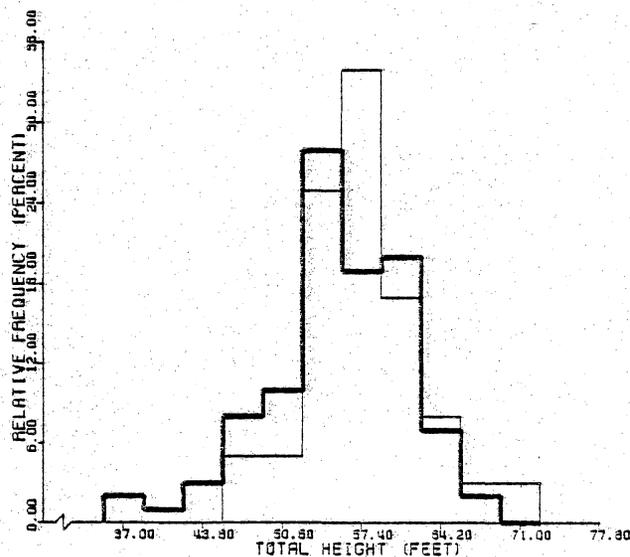


b. m deterministic  
height stochastic

Figure 7. Observed height distribution at age 21 on plot 41 of the Hill Farm data, and height distributions predicted with the parameter distribution model.



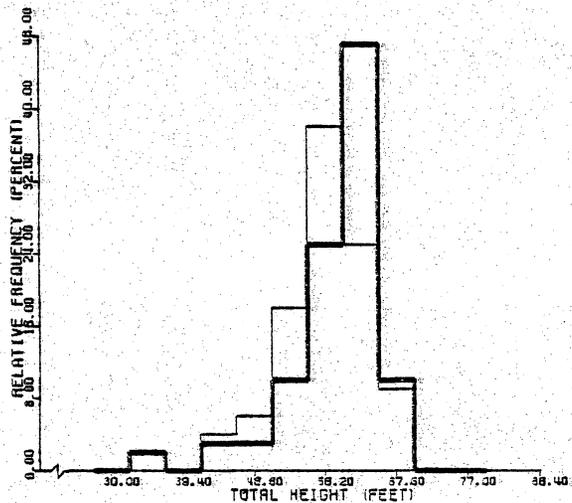
a. m deterministic  
 height deterministic



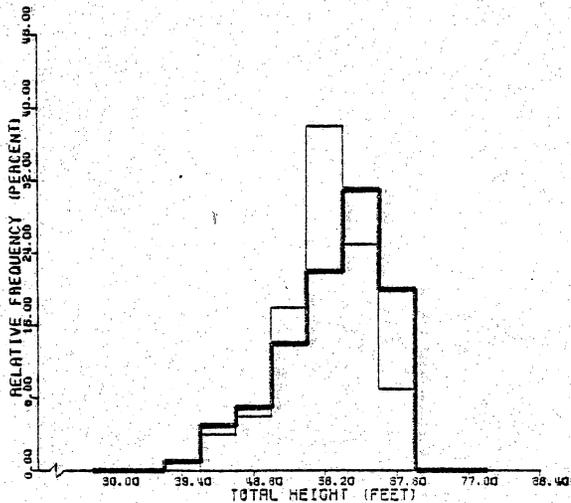
b. m deterministic  
 height stochastic

Figure 8. Observed height distribution at age 21 on plot 42 of the Hill Farm data, and height distributions predicted with the parameter distribution model.

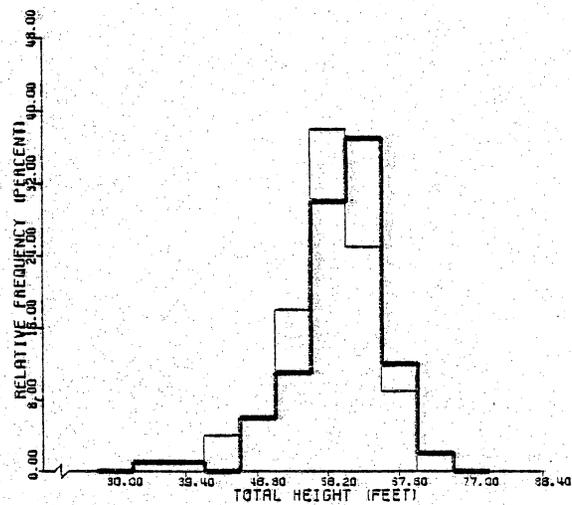
— OBSERVED  
 — PREDICTED



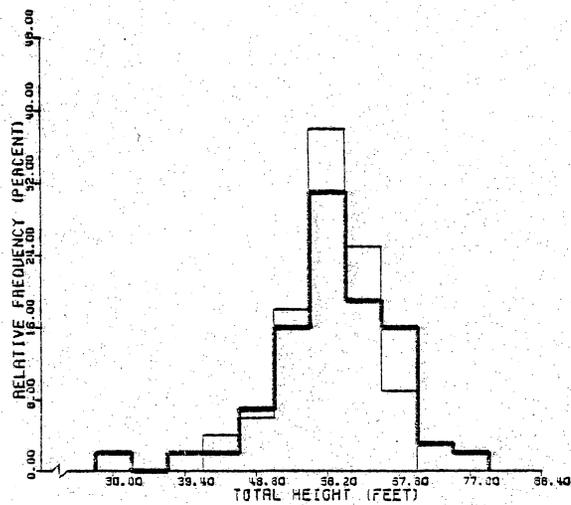
a. m deterministic  
 height deterministic



b. m deterministic  
 height stochastic



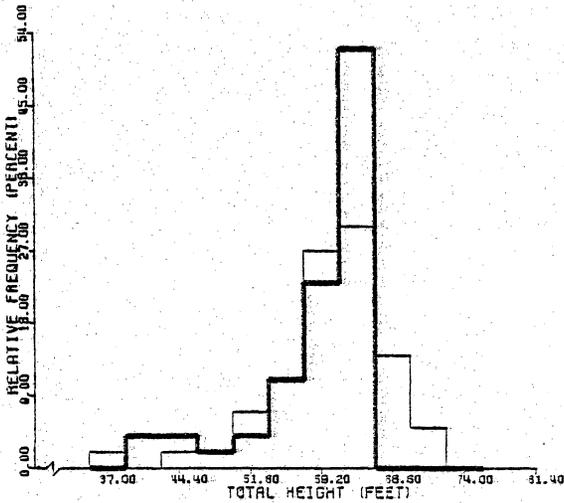
c. m stochastic  
 height deterministic



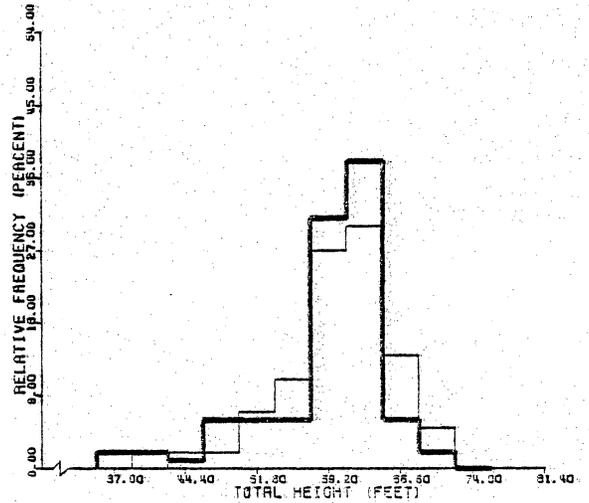
d. m stochastic  
 height stochastic

Figure 9. Observed height distribution at age 20 on plot 43 of the Hill Farm data, and height distributions predicted with the parameter distribution model.

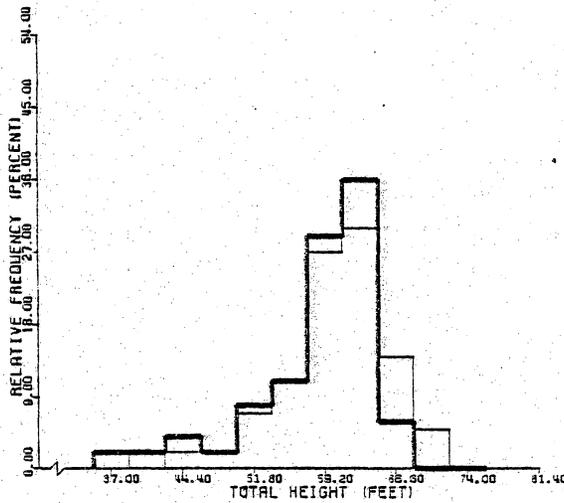
— OBSERVED  
 — PREDICTED



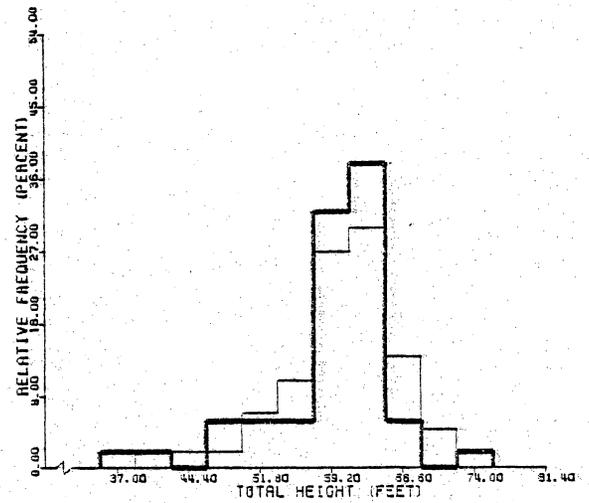
a.  $\mu$  deterministic  
 height deterministic



b.  $\mu$  deterministic  
 height stochastic



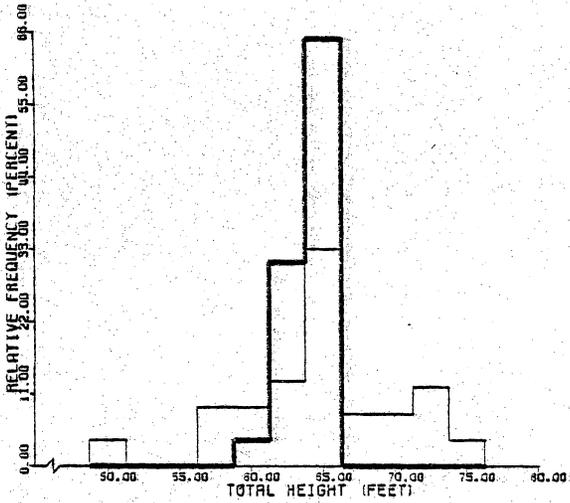
c.  $\mu$  stochastic  
 height deterministic



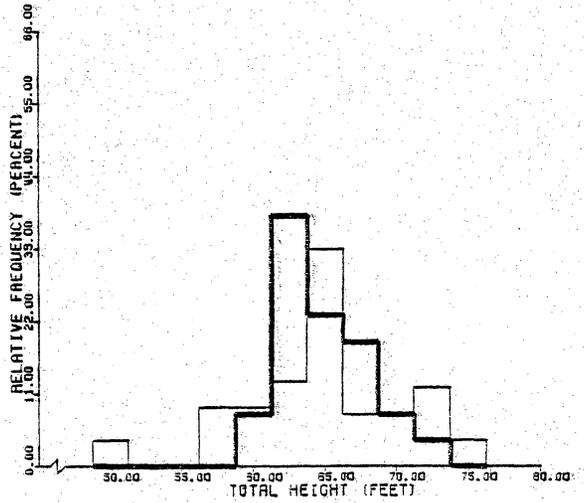
d.  $\mu$  stochastic  
 height stochastic

Figure 10. Observed height distribution at age 20 on plot 44 of the Hill Farm data, and height distributions predicted with the parameter distribution model.

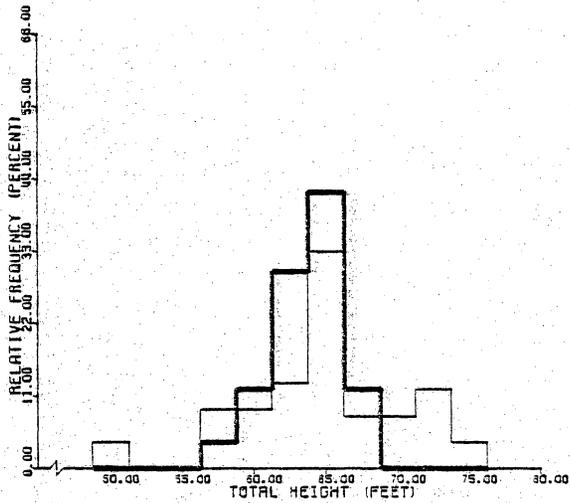
— OBSERVED  
 — PREDICTED



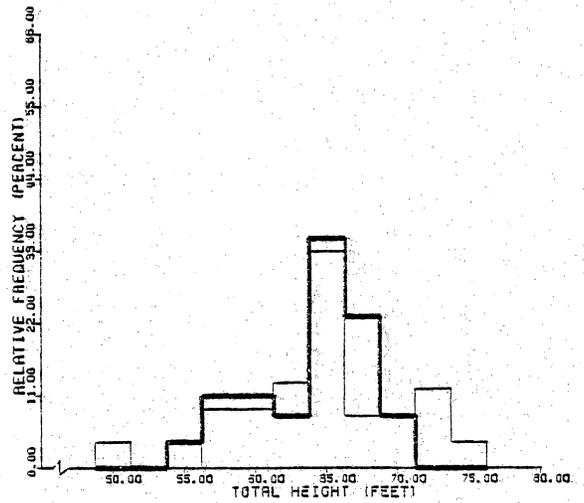
a. m deterministic  
 height deterministic



b. m deterministic  
 height stochastic



c. m stochastic  
 height deterministic



d. m stochastic  
 height stochastic

Figure 11. Observed height distribution at age 21 on plot 45 of the Hill Farm data, and height distributions predicted with the parameter distribution model.

15. The results obtained with this seed were typical of the results obtained from runs with 19 different seeds. Table 16 shows the means and standard deviations of the predicted means and predicted standard deviations of basal area resulting from all 20 runs of the model with 20 different seeds. Table 17 shows the same for heights.

An internal difficulty in the model may be evident in Figure 2, which shows the predicted and observed basal area distributions for plot 41. Recall that the marginal distributions of  $k$  were fit to the coefficients estimated for every tree with at least three measurements on each plot. Thus trees which did not survive to the last measurement were included in this fitting data set for the marginals. The author sees no way to circumvent this difficulty. No matter what age is picked, if only stems which survived to that age are used to calibrate the internal models, the marginals of  $k$  will still be biased as described above for projected ages. They would also be biased prior to that age as stems which had not yet died would not be allowed to influence the marginal of  $k$ . On plots thinned to 1000 TPA, average mortality was approximately 40%. In Figure 2 it is evident that the model predicted too many stems with low basal areas for this plot. Presumably this was due to the inclusion in the fitting set for the marginals of  $k$  many

stems on plots 11, 21, and 31 which had died prior to the last measurement. These stems would logically have been mostly smaller stems, resulting in a predicted distribution too heavy at the lower end. Despite this problem, the predicted and observed distributions in Figure 2 do not appear to be too radically different from a practical standpoint. It also must be kept in mind that 1000 TPA at age 5 is very heavy stocking by current standards. Although the figures vary, plantations are now more commonly established in the 500-600 TPA range. When comparing the observed and predicted distributions for plots 42-45 (Figures 3-6) this problem was not evident. Note further that the difficulty caused by mortality is not unique to this model. In diameter distribution models if one attempts to establish a biological basis for the p.d.f. employed by solving for the implied growth function as Bailey (1980) has, then one must assume that either no mortality occurs or that mortality is proportional across the diameter distribution. The limitations of either of these assumptions are obvious.

Figures 3-6 show that with the exception of Figures 5a and 6a, the predicted and observed basal area distributions generally corresponded well for plots 42-45. There are particular cells in given histograms where the two diverge

widely, but this is not surprising since the two distributions are both random variables. With larger sample sizes, i.e., larger plots, these differences would be expected to become minor.

Figures 4-6 reveal why allowing  $m$  to be stochastic tends to improve the correspondence between observed and predicted distributions. This method produced a wider spread in the predicted distribution, which agreed more closely with the observed.

Table 16 shows that the means of the predicted distributions tended to be close to those of the observed distributions. Only on plot 42 (for the single prediction system used for this plot) and plot 44 (for both systems) were the observed means more than one standard deviation away from the predicted means. However, the differences between the observed and predicted means for plot 42 and the two predicted means on plot 44 were only approximately 18, 12, and 12 percent of the observed means, respectively. The author feels that these differences are not excessively large. The reason for the over-predictions on these two plots is evident after consideration of the means on the plots used for fitting. Observed mean basal areas at the last measurement on plots 12, 22, and 32 were .3371, .3913, and .2684 square feet, respectively. For plots 14, 24, and

34 the means were .6497, .6154, and .6423 square feet respectively. The predicted means for plots 42 (.3261 sq. ft.) and 44 (.6375 sq. ft., m deterministic; .6371 sq. ft., m stochastic) agree well with these values.

Table 16 also shows that on the average, the mean of the predicted distribution was slightly higher than the observed mean, suggesting that mortality was not a problem in this regard. The mean absolute difference between observed and predicted means was .03 sq. ft. for both deterministic and stochastic m. The mean absolute difference between observed and predicted standard deviations was .05 sq. ft. for deterministic m and .06 sq. ft. for stochastic m. As would be expected, allowing m to be stochastic resulted in higher standard deviations of predicted basal area on plots where this method was used. On plots 43 and 44 the standard deviation with m deterministic was closest to the observed, whereas on plot 45 the standard deviation with m stochastic was closest. This is in contrast to the p-levels and histograms which show that allowing m to be stochastic improves the correspondence between the observed and predicted distributions. Although the p-levels and histograms were from one run of the model, the above trend in standard deviations was also evident in this one run (see Table 14). Thus the correspondence

between the observed and predicted distributions depends on more than correspondence between the standard deviations.

Figures 7 through 11 depict the observed and predicted height distributions. An ocular inspection shows that the two distributions tend to take the same shape. When there is an obviously poor correspondence, this may be traced back to the predicted basal area distribution. For example, Figure 7 shows that on plot 41, both prediction methods (m deterministic, H deterministic; m deterministic, H stochastic) resulted in too many low heights. The reason for this was probably the previously discussed under-prediction of basal areas on this plot. Another example is Figure 11a. In this case the predicted heights are concentrated in too small an interval. Figure 6a shows that this was also true of the predicted basal areas.

The values in Table 17 verify the conclusions drawn from the histograms. The average absolute differences between observed and predicted mean heights were 2.72 ft., 2.72 ft., 1.56 ft., and 1.53 ft. for prediction methods I, II, III, and IV respectively. The first two average absolute differences are larger than the second two because methods I and II were used on plot 41 and methods III and IV were not. The differences between observed and predicted mean heights (observed - predicted) for plot 41 were 7.36

ft. and 7.4 ft. for methods I and II, respectively. These under-predictions were due to the over-abundance of low heights in the predicted distributions.

The average absolute differences between observed and predicted standard deviations of height were 1.92 ft., 1.73 ft., 0.90 ft., and 1.13 ft. for methods I, II, III, and IV, respectively. Again the reason the first two average absolute differences are larger than the second two is the inclusion of the standard deviations from plot 41 in the first two. The differences between observed and predicted standard deviations (observed - predicted) were 4.88 ft. and 5.35 ft. for prediction methods I and II, respectively. These over-predictions were due to the over-abundance of low heights in the predicted distributions. On plot 45, prediction method I underestimated the observed standard deviation by 3.74 ft. This was more than twice as large as the differences yielded by the other methods. The reason for the under-prediction of the standard deviation by method I was the concentration of the predicted basal areas into too small an interval on this plot when  $m$  was deterministic, as described earlier.

It is interesting that method IV ( $m$  stochastic,  $H$  stochastic) was better on only one plot, 45. On this plot the three levels of randomness ( $k$ ,  $m$ , and  $H$  all stochastic)

resulted in a higher standard deviation which more closely agreed with the observed. On plots 43 and 44 this method produced too great a standard deviation.

The performance of the survival curve can be evaluated by comparing the predicted and observed numbers of stems at the last measurement for each plot. These values are shown in Table 14. Only on plot 21 were the observed and predicted numbers of trees substantially different. In comparison with plots 11, 31, and 41 it is apparent that plot 21 experienced heavier than average mortality. Note that the observed and predicted number of trees in Table 15 cannot be compared because not every observed tree had a recorded height, yet a height was predicted for every predicted basal area.

### 5.3 NONLINEAR GROWTH FUNCTION

Fitting the nonlinear growth function with the asymptote specified at 22.3 square feet of basal area resulted in the following estimated regression equation:

$$BA_{i+1} = A(1 - e^{-f_i(AP_{i+1}-c)})g_i$$

where:  $f_i = \{(0.00400 CR_i + 6.15593 BA_i - 0.00010 TS_i - 0.00237 B_i + 0.00499 APA_i)/1000\}$

$$g_i = 0.44272 - 0.02351 TS_i/1000$$

$$R^2 = 0.93 \quad Sy.x = 0.048 \quad n = 10195 \quad \text{mean } BA_{i+1} = 0.1927$$

where:  $BA_{i+1}$  = basal area of stem at time  $i+1$  in square feet

$CR_i$  = predicted crown ratio of stem at time  $i$

$B_i$  = basal area per acre (sq. ft.) at time  $i$

$TS_i$  = number of trees per acre at time  $i$

$APA_i$  = area potentially available to the stem at time  $i$

$AP_i$  = age form planting at time  $i$ .

None of the asymptotic 95% confidence intervals for the estimated coefficients included zero. The  $R^2$ ,  $Sy.x$ , and estimated coefficients were all very similar when the asymptote was specified at 25.65 sq. ft. or 18.96 sq. ft.

Note that the estimated coefficient for  $TS$  in the function modelling the quantity  $(1/(1-m))$  was negative. This means that lower  $m$  values were predicted for higher densities, which is consistent with the results from the parameter distribution model.

The estimated linear model was

$$BA_{i+1} = -0.0896 + 0.0010 CR_i + 0.0003 APA_i + 1.0694 BA_i - 0.0001 B_i + (3.19 \times 10^{-6}) TS_i + 0.0216 (AP_{i+1} - AP_i)$$

$$R^2 = 0.95 \quad Sy.x = 0.040 \quad n = 10195 \quad \text{mean } BA_{i+1} = 0.1927$$

The estimated coefficient for  $TS_i$  was significant at the .1 level of probability. All the others were significant at the .01 level.

Results of fitting the non-linear and linear models on the independent data are presented in Table 18. The Table displays the mean difference and mean absolute difference resulting from predicting the observed basal areas on replication four, and then computing the difference between the observed and predicted. The values show that the linear model performed better than the nonlinear.

The standard error of the mean difference between observed and predicted basal areas for the nonlinear model, asymptote = 18.96 sq. ft. was  $.01 \times 10^{-4}$  larger than those for the nonlinear model with asymptotes of 22.3 sq. ft. or 25.65 sq. ft. The three nonlinear models were identical under the three other statistics shown in Table 18. This again establishes the insensitivity of the C-R curve to the asymptote.

Table 18. Means and standard errors of means of difference and absolute difference between basal area observed on one replication in the Hill Farm data, and basal area predicted with the linear model and with the nonlinear model with three different asymptotes.\*

Model	n	Difference		Absolute Difference	
		mean	std. error of mean	mean	std. error of mean
linear	3391	0.0006	$7.19 \times 10^{-4}$	0.0253	$5.74 \times 10^{-4}$
nonlinear, I	3391	-0.0071	$8.51 \times 10^{-4}$	0.0302	$6.87 \times 10^{-4}$
nonlinear, II	3391	-0.0071	$8.51 \times 10^{-4}$	0.0302	$6.87 \times 10^{-4}$
nonlinear, III	3391	-0.0071	$8.52 \times 10^{-4}$	0.0302	$6.87 \times 10^{-4}$

\* nonlinear, I : asymptote = 22.3 square feet.  
 nonlinear, II : asymptote = 25.65 square feet.  
 nonlinear, III : asymptote = 18.96 square feet.

## Chapter 6

### CONCLUSIONS AND RECOMMENDATIONS

The data set used in this study had little site quality variation, and a specific combination spacing-selection thinning. Therefore, care must be exercised in extrapolating the results found herein to cases where site quality is markedly different from that of the data used here, or the thinning procedure is different.

As stated earlier, the model

$$H = a + b \log(BA)$$

performed well in this study. Based on these preliminary findings, it is suggested that a larger study is needed in order to determine whether this model is better in general than other candidate height - diameter/basal area curves.

Although a much larger study of the Parameter Distribution model is needed in order to draw any firm conclusions, it appears now that the following general procedure would be the best means of employing this model. First be cautious when applying the model on plots which were established at high densities or have experienced heavy mortality. Second, test to see at what densities  $m$  may be set to zero, and do so for these densities. For other densities regress  $m$  on  $k$ ,  $k^2$ , and stand variables measured

at a reference point in time. For prediction of basal areas employ the stochastic  $m$ . This method appeared to function better than using the deterministic  $m$  on plots where  $m$  was determined through regression analysis. For prediction of heights, use the deterministic height model. Based on Table 16 this method appeared to perform as well as the others.

In retrospect, the Parameter Distribution model has at least five desirable characteristics. First, it is based on a sound biological growth curve. Second, it possesses an internal stochastic element. Third, all the internal models appear to fit well, in contrast to the diameter distribution models in which the internal prediction equations for the parameters of the p.d.f. tend to fit the observed data poorly. Fourth, again in contrast to the diameter distribution models, the predicted basal area distribution is not constrained to conform to a certain p.d.f. Finally, and perhaps most importantly, the model is logical and relatively easy to comprehend. However, the parameter distribution model also has at least two disadvantages. First is the problem of mortality. More precisely, the problem is the death of trees included in the fitting data set for the marginal distribution of  $k$  before they reach projection age. Although this did not appear to have a noticeable effect in the range of densities of practical

interest, more studies are needed to fully define the extent of this problem. As stated earlier, it does not appear to the author that this problem can be eliminated. However, if necessary, perhaps a method could be devised to alleviate the effects of this problem. The objective of such a study might be to determine the optimum minimal lifespan for trees included in the fitting data. This might be found by varying the minimal lifespan and comparing the correspondence of the resulting predicted basal area distributions to the observed.

The second disadvantage of the Parameter Distribution model is that it requires re-measurement data to calibrate the growth functions. In situations where only temporary plot data are available, other techniques must be applied. The good fit of the C-R curve to the individual stems considered in this study suggests that some use should be made of such a function in diameter distribution models. Some work has progressed along these lines. As previously mentioned, Bailey (1980) has derived the growth functions implied by the use of several p.d.f.'s commonly used in diameter distribution models. These growth curves should be examined to see if they represent individual tree growth adequately. If they do not, then an effort should be made to find or develop a p.d.f. which would include a curve such

as the C-R growth curve as the implied growth function. Perhaps a method could be derived to incorporate such a growth curve into the estimation of the parameters of the diameter or basal area distribution. For example, perhaps the method of solving for the parameters of the distribution from predicted quantities which are functions of the moments of the p.d.f. could be carried a step further in which a sound growth curve is employed to estimate the moments.

Regarding the individual tree nonlinear growth function, on the basis of the results reported herein, it does not yield better growth predictions than the simpler linear model. Therefore the use of the nonlinear model would ordinarily be uncalled for. However, foresters are often called on to extrapolate beyond the range of observed data. For this purpose the nonlinear model would probably be safer than the linear model, since it is based upon a reasonable growth curve with known long-term properties.

In summary, the results of this study were mixed. The effort devoted to height-diameter curves indicated that the model currently in widespread use may not be the best available model. The Parameter Distribution model, which estimates the basal area and height distributions at given points in time performed fairly well. More work is needed to refine this model and link it more closely with

alternative methodologies. The nonlinear individual tree growth function did not perform as well as an alternative linear model, and its use would only be justified when extrapolation is necessary.

## LITERATURE CITED

- Alder, D. 1979. A distance-independent tree model for exotic conifer plantations in east Africa. *For. Sci.* 25:59-71.
- Alemdag, I.S. 1978. Evaluation of some competition indexes for the prediction of diameter increment in planted white spruce. *For. Man. Inst., Ottawa, Ont., Inf. Rep. FMR-X-108.*
- Arney, J.D. 1972. Computer simulation of Douglas-fir tree and stand growth. Unpublished Ph.D. Thesis, Oregon State Univ., 79 p.
- Assman, E. 1970. The principles of forest yield study. 1st Eng. Ed., Pergamon Press, Ltd. 506 p.
- Bailey, R.L. 1980. Individual tree growth derived from diameter distribution models. *For. Sci.* 26: 626-32.
- Bailey, R.L. and T.R. Dell, 1973. Quantifying diameter distributions with the Weibull function. *For. Sci.* 19: 97-104.
- Beck, D.E. 1971. Height-growth patterns and site index of white pine in the southern Appalachians. *For. Sci.* 17: 252-60.
- Bella, I.E. 1970. Simulation of growth, yield and management of aspen. Unpublished Ph.D. Thesis, Univ. of British Columbia, 190 p.
- Bickel P.J. and K.A. Doksum. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, Inc. San Francisco, Calif. 493 p.
- Bruner, H.D. and J.W. Moser, Jr. 1973. A markov chain approach to the prediction of diameter distributions in uneven-aged stands. *Can. J. For. Res.* 3: 409-17.
- Buongiorno, J. and B.R. Michie. 1980. A matrix model of uneven-aged forest management. *For. Sci.* 26: 609-25.
- Burkhart, H.E. and R.B. Tennent. 1977. Site index equations for radiata pine in New Zealand. *N. Zea. J. For. Sci.* 7: 408-16.

- Burkhart, H.E. and M.R. Strub. 1974. A model for simulation of planted loblolly pine stands. In Growth Models for Tree and Stand Simulation (Ed. J. Fries). Royal Coll. of Forestry, Stockholm, Sweden, pp. 128-35.
- Burkhart, H.E., Q.V. Cao, and K.D. Ware. 1981. A comparison of growth and yield prediction models for loblolly pine. VPI and SU, FWS-2-81, 59 p.
- Carmean, W.H. 1972. Site index curves for upland oaks in the Central States. For. Sci. 18: 109-20.
- Chapman, D.G. 1961. Statistical problems in population dynamics. Proc. Fourth Berkeley Symp. Math. Stat. and Prob., Univ. Calif. Press, Berkeley and Los Angeles. pp. 153-68.
- Clutter, J.L. and F.A. Bennett. 1965. Diameter distributions in old-field slash pine plantations. Ga. For. Res. Council. Rep. 13, 9 p.
- Curtis, R.O. 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. For. Sci. 13: 365-75.
- Dale, M.E. 1975. Individual tree growth and simulation of stand development of an 80-year-old white oak stand. In Forest Modeling and Inventory (A.R. Ek, J.W. Balsiger and L.C. Promnitz. Eds.) Dept. of For. Univ. of Wisc. pp. 49-63.
- Daniels, R.F. 1981. An integrated system of stand models for loblolly pine. Unpublished Ph.D. Thesis. VPI and SU. 101 p.
- Daniels, R.F. and H.E. Burkhart. 1975. Simulation of individual tree growth and stand development in managed loblolly pine plantations. VPI and SU, FWS-5-75, 69 p.
- Daniels, R.F., H.E. Burkhart, G.D. Spittle, and G.L. Somers. 1979. Methods for modelling individual tree growth and stand development in seeded loblolly pine stands. VPI and SU, FWS-1-79, 50 p.
- Ek, A.R. 1973. Performance of regression models for tree height estimation with small sample sizes. In Statistics in Forestry Research, Proc. I.U.F.R.O. subject group S6.02, Vancouver, Can., Aug. 1973. Ed. W.G. Warren. pp 67-80.

- Ek, A.R. and R.A. Monserud. 1974. FOREST: A computer model for simulating the growth and reproduction of mixed species forest stands. Univ. Wisc. School of Nat. Res. Res. Pap. R2635. 13 p.
- Errikson, L. 1978. Single-tree competition models, predicting stand development after cleaning, in: Growth models for long-term forecasting of timber yields, Fries, et. al. ed., VPI & SU, FWS-1-78, 249 p.
- Feduccia, D.P., T.R. Dell, W.F. Mann, T.E. Campbell and B.H. Polmer. 1979. Yields of unthinned loblolly pine plantations on cutover sites in the West Gulf Region. U.S.D.A. For. Serv. Res. Pap. SO-148, 88 p.
- Furnival, G.M. 1961. An index for comparing equations used in constructing volume tables. For. Sci. 7: 337-41.
- García, O. 1979. Modelling stand development with stochastic differential equations. In F.R.I. Symposium No. 20 "Mensuration for Management Planning of Exotic Forest Plantations". New Zealand For. Serv., For. Res. Inst. compiled by D.A. Elliott. pp. 315-33.
- Goldstein, H.B. 1979. The Design and Analysis of Longitudinal Studies; Their Role in the Measurement of Change. Academic Press, New York and London, 199 p.
- Goulding, C.J. 1972. Simulation techniques for a stochastic model of the growth of Douglas-fir. Ph.D. Thesis, Univ. of British Columbia. 234 pp.
- Graney, D.L. and H.E. Burkhart. 1973. Polymorphic site index curves for shortleaf pine in the Ouachita Mountains. U.S.D.A. For. Serv. Res. Pap. SO-85. 12 p.
- Greenhill, A.G. 1881. Determination of the greatest height consistent with stability that a vertical pole or mast can be made, and of the greatest height to which a tree of given proportions can grow. Proc. Camb. Phil. Soc. IV, Part II, pp. 65-73.
- Grosenbaugh, L.R. 1965. Generalization and reparameterization of some sigmoid and other nonlinear functions. Biometrics 21: 708-14.
- Guttenberg, R. Von. 1915. Wachstum and Ertrag der Fichte im Hochgebirge. Wien und Leipzig. 153 p.

- Hafley, W.L. and H.T. Schreuder. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* 7: 481-87.
- Hahn, J.T. and R.A. Leary. 1979. Potential diameter growth functions. In *A generalized forest growth projection system applied to the Lake States Region*. U.S.D.A. For. Serv. Gen. Tech. Rep. NC-49. pp 22-26.
- Hatch, C.R. 1971. Simulation of an even-aged red pine stand on northern Minnesota. Unpublished Ph.D. Thesis. Univ. Minn. 120 p.
- Henriksen, H.A. 1950. Height-diameter curve with logarithmic diameter. *Dansk Skovforen. Tidsskr.* 35: 193-202. (in Danish; English summary in *For. Abstr.* 13: 198.)
- Hoel, P.G., S.C. Port and C.J. Stone. 1972. *Introduction to Stochastic Processes*. Houghton Mifflin Co. Boston. 203 p.
- Hyink, D.M. Diameter distribution approaches to growth and yield modelling. In *Forecasting Forest Stand Dynamics*. Lakehead Univ., Thunder Bay, Ontario, Canada (in press).
- Jackson, D.S. and J. Ure. 1964. The control of basal area increment in young *Pinus radiata*. *N. Zea. J. For.* 9: 78-88.
- Karish, J.F. 1973. Parameter correlation effects in non-linear mathematical models for biological growth. Unpublished M.S. Thesis. Penna. State Univ. 68 p.
- Krause, G.F., P.B. Siegel, and D.C. Hurst. 1967. A probability structure for growth curves. *Biometrics* 23: 217-25.
- Lenhart, J.D. 1972. Cubic volume yields for unthinned old-field loblolly pine plantations in the Interior West Gulf Coastal Plain. Stephen F. Austin State Univ., Texas For. Pap. 14, 46 p.
- Lenhart, J.D. and J.L. Clutter. 1971. Cubic-foot yield tables for old-field loblolly pine plantations in the Georgia Piedmont. *Ga. For. Res. Counc. Rep.* 22 - Series 3, 12 p.
- Lohrey, R.E. 1974. Growth and yield in a longleaf pine plantation, in: *Symposium on management of young pines*, Alexandria, La. and Charleston, S.C., U.S.D.A. For. Serv. pp. 88-96.

- Matney, T.G. and A.D. Sullivan. Compatible stand and stock tables for thinned and unthinned loblolly pine stands. For. Sci. (in press).
- Meyer, H.A. 1940. A mathematical expression for height curves. J. For. 38: 415-20.
- Michailoff, I. 1943. Zahlenmässiges Verfahren für die Ausführung der Bestandeshöhenkurven. Forstw. Cbl. u. Thar. Forstl. Jahrb. 6: 273-79.
- Mitchell, K.J. 1975. Dynamics and simulated yield of Douglas-fir. For. Sci. Mono. 17. 39 p.
- Moore, J.A., C.A. Budelsky and R.C. Schlesinger. 1973. A new index representing individual tree competitive status. Can. J. For. Res. 3: 495-500.
- Munro, D.D. 1974. Forest growth models--a prognosis. In Growth Models for Tree and Stand Simulation (Ed. J. Fries). Royal Coll. of Forestry. Stockholm, Sweden. pp. 7-21.
- Moser, J.W., Jr. 1967. Growth and yield models for uneven-aged forest stands. Ph.D. thesis. Purdue Univ. 149 p.
- Newnham, R.M. and H. Mucha. 1971. A test of the effectiveness of different competition indices in predicting diameter growth in a young red pine stand. For. Man. Inst., Ottawa, Ont., Int. Rep. FMR-20. 30 p.
- Newnham, R.G. and J.H.G. Smith. 1964. Development and testing of stand models for Douglas-fir and lodgepole pine. Forestry Chron. 40: 492-502.
- Pardo, R. 1973. A.F.A.'s social register of big trees. Amer. For. 79: 21-47.
- Peden, L.M., J.S. Williams, and W.E. Frayer. 1973. A Markov model for stand projection. For. Sci. 19: 303-14.
- Pienaar, L.V. 1965. Quantitative theory of forest growth. Ph.D. thesis, Univ. of Wash. 177 p.
- Pienaar, L.V. and K.J. Turnbull. 1973. The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. For. Sci. 19: 2-22.

- Prodan, M. 1968. Forest Biometrics. Pergamon Press, Oxford. 447 p.
- Rao, C.R. 1965. The theory of least squares when the parameters are stochastic and its application to growth curves. *Biometrika* 52, 3 and 4, 447-58.
- Richards, F.J. 1959. A flexible growth function for empirical use. *J. Exp. Bot.* 10: 290-300.
- \_\_\_\_\_. 1969. The quantitative analysis of growth. in: *Plant Physiology, A Treatise*, ed. F.C. Steward, vol. VA. Academic Press, New York and London.
- Rose, D.W. and C.M. Chen. 1977. Nonlinear biological yield models for jack pine. *Minn. For. Res. Notes*, No. 262. 4 p.
- SAS Institute. 1979. SAS User's Guide. 1979 Ed. SAS Inst. Inc., Raleigh, N.C.
- Schreuder, H.T. and W.L. Hafley. 1977. A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* 33: 471-78.
- Smalley, G.W. and R.L. Bailey. 1974. Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama, and Georgia highlands. U.S.D.A. For. Serv. Res. Pap. SO-96. 81 p.
- Stage, A.R. 1973. Prognosis model for stand development. U.S.D.A. For. Serv. Res. Pap. INT-137. 32 p.
- \_\_\_\_\_. 1975. Prediction of height increment for models of forest growth. U.S.D.A. For. Serv. Res. Pap. INT-164. 20 p.
- Stoffels, A. and J. Van Soest. 1953. The main problems in sample plots. 3. Height regression. *Ned. Boschb. Tijdschr.* 25: 190-99. [In Dutch; English summary in *For. Abstr.* 15; 77.]
- Stout, B.B. 1973. Height-diameter relations in trees: an examination of Greenhill's model. In *I.U.F.R.O. Biomass Studies, Proc. of Joint Meeting of S4.01 and S6.02*, Aug 20-24, 1973. Vancouver, B.C., Can. pp. 159-70.
- Trorey, L.G. 1932. A mathematical method for the construction of diameter height curves based on site. *For. Chron.* 8: 121-32.

Turnbull, K.J. 1963. Population dynamics in mixed forest stands. A system of of mathematical models of mixed stand growth and structure. Ph.D. thesis. Univ. of Wash. 187 p.

Von Bertalanffy, L. 1957. Quantitative laws in metabolism and growth. The Quart. Rev. of Bio. 32: 217-31.

**The vita has been removed from  
the scanned document**

MODELS OF STAND BASAL AREA DISTRIBUTIONS, INDIVIDUAL TREE  
BASAL AREA GROWTH,  
AND HEIGHT-DIAMETER RELATIONSHIPS FOR LOBLOLLY PINE

by

Edwin James Green

(ABSTRACT)

The study dealt with developing methodologies for predicting basal area distributions and individual tree basal areas. Data for the study was from the Hill Farm Experiment Station at Homer, Louisiana.

Five height-diameter (basal area) curves were examined to determine which was most appropriate for the data set utilized. The model  $H = a + b \log(BA)$ , where  $H$  denotes height and  $BA$  denotes basal area, was chosen as best, based on several fit and prediction oriented statistics.

A stochastic basal area distribution model, called the parameter distribution model, was developed. The model was based on the Chapman-Richards growth curve. This curve was fit to all stems on approximately 3/4 of the data set. Two parameters of the curve were fixed a priori, leaving two parameters to be estimated. A sampling distribution was fit to the estimates of the rate parameter,  $k$ . Models were developed to predict the parameters of this distribution

from stand variables. A model was then derived to predict  $m$ , the shape parameter of the C-R curve, from  $k$  and stand variables. Finally, an existing survival function was modified. The overall model was implemented as follows: first, the number of surviving stems was predicted. Then  $k$  and  $m$  values were predicted for each predicted stem. Substitution of these two values into the C-R curve yielded a predicted basal area for each stem. The previously mentioned height-diameter curve was employed to predict a height for each predicted basal area. Stochastic elements were built into the prediction model for  $m$  and the height-diameter curve. Predicted basal area and height distributions were compared to observed on the remaining 1/4 of the data set. Although the two-sample K-S test was statistically significant, the observed and predicted distributions did appear to be close, in general, from a practical standpoint. This approach appears promising as a stochastic method of predicting size distributions.

The Chapman-Richards curve was also modified for use as an individual tree basal area growth model. Two parameters of the curve were fixed, and the remaining two were modelled as functions of tree- and stand-level variables. The modified growth function fit the data well, but on an independent data set, a simpler linear model of basal area growth

performed better in terms of mean difference and mean absolute difference between observed and predicted basal areas. Thus, the only anticipated use of the modified C-R model is in situations where extrapolation beyond the range of observed data is required, since this model has desirable long-term characteristics, whereas the linear model does not.