An Examination of Methods for Localizing Site Index Equations

by

David Kevin Walters

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Forestry

APPROVED:

Dr. Harold E. Burkhart, Chairman

Dr. Timothy G. Gregoire

Dr. Marion R. Reynolds

June 1987

Blacksburg, Virginia

An Examination of Methods for Localizing Site Index Equations

by

David Kevin Walters

Dr. Harold E. Burkhart, Chairman

Forestry

(ABSTRACT)

Equations for modeling the height-age pattern of forest trees or stands are typically developed for a given species in a specified region. In order to adequately model height-age patterns, the resulting equations are often quite complex. This study addresses the question of how to increase the accuracy and precision of the prediction of height-age relationships through the use of localized equations.

Although the pattern of height-age relationships of loblolly pine plantations was studied, the methodology should prove valuable for other species as well. The Schumacher logarithm of height-reciprical of age model was fitted to data from loblolly pine plantations to attain an average guide curve. Various methods of localizing this equation to a particular stand were examined. The methods are based on empirical Bayesian, maximum likelihood, and Kalman filter theory. All of these methods employ the general concept of feedback in localizing the simple equation. The best of these various models is compared with the unadjusted model and a more complex polymorphic equation. The adjusted model compares favorably with these other two models.

Acknowledgements

For their comments and suggestions, I wish to thank Drs. Timothy Gregoire and Marion Reynolds. I am especially indebted to Dr. Harold Burkhart for his assistance with this project.

The data and financial support for this study were provided by the Loblolly Pine Growth and Yield Research Cooperative at VPI & SU. Special thanks are extended to the supporters of this Cooperative: Bowater Incorporated, Champion International Corporation, Chesapeake Corporation, Federal Paper Board Company, International Paper Company, Union Camp Corporation, Westvaco Corporation, Weyerhaeuser Company, North Carolina Division of Forestry, Virginia Division of Forestry, and the U.S. Forest Service.

In addition, I wish to thank all my friends for providing me with invaluable moral support. To Karen Bessling, Mary Bleier, and Michael Dyer, I am especially grateful. The actual preparation of this manuscript would have been extremely difficult to complete without the assistance of Cathy Barker. Her efforts are deeply appreciated. My fiancee and colleague, Terese Walsh, played a special role in the preparation of this manuscript. Her assistance, both technically and in simply encouraging me to take an occasional break, is deeply appreciated.

Finally, I wish to thank my parents for their support and encouragement during my years in school.

Table of Contents

List of Tables

List of Illustrations

Chapter I

INTRODUCTION AND JUSTIFICATION

The site quality of forest land has been evaluated in many ways. The most common and well accepted approach uses dominant stand height at a given (index) age, referred to as site index, as a measure of the quality of the site. Consequently, a large number of studies have developed equations for estimating site index for a variety of species and locales. These equations estimate the height-age (H/A) pattern of a stand and the site index is obtained from these H/A curves. In addition, a variety of equation types and methods of fitting these equations have been developed. The development of H/A curves has progressed from graphical techniques (Bruce, 1926) to statistical methods (regression analysis being by far the most prevalent procedure). Early works utilized simple equations such as:

$$\log(H) = a_0 + a_1 A^{-1} \qquad [1.0]$$

to portray height development (Schumacher, 1939). This equation has often been used for a "guide curve" to produce an anamorphic (or harmonized) family of H/A curves.

Later works utilized more complex equations and polymorphic site index curves. Site index equations have become increasingly complex in an effort to adequately describe the wide variety of height-age patterns that are observed. An alternative to developing increasingly complex equation forms is to utilize a relatively simple equation, such as Equation [1.0], and try to "localize" the equation to a specific stand. Several methods of localization or adjustment lend themselves quite readily to this problem.

This study addresses the questions of how to increase the accuracy and precision of site index prediction by localizing relatively simple equations. Site index prediction is usually divided into two separate phases: (1) research and development of H/A curves, and (2) application of H/A curves to estimate site index.

The general objective of this study was to examine methodologies for refining or adjusting site index curves before they are applied. As previously stated, a set of H/A curves is typically developed for a species in a general region. These curves are then used in a specific stand by sampling the stand for height and age. As an alternative, the sample information may first be used to adjust the H/A curves to the particular stand. Specifically this adjustment may:

1. Allow for more accurate and precise estimation of the H/A pattern in a particular stand.
2. Decrease the need for developing increasingly complex equations.
3. Decrease the need for developing separate localized equations.
4. Increase the accuracy and precision of estimates of the H/A development when the input data are from young stands.

Chapter II

LITERATURE REVIEW

The intensive silvicultural techniques applied to forests and the importance of economically efficient management make the estimation of site quality possibly more important now than at any previous time. Ideally, a single numerical estimate of the productive capacity of the site (site quality) is desired. This estimate has previously been made both indirectly and directly. To be useful, any estimator of site quality must be related to the product to be harvested from the site. Clutter et al. (1983) provide an apt definition of site quality within the context of timber management: "the timber production potential of a site for a particular species or forest type".

Indirect methods can be used to estimate site quality from overstory interspecies relationships, understory characteristics, geographic, topographic, and climatic factors. This list is by no means exhaustive. Direct estimation of site quality can be made from previous stands occurring on a site, periodic height growth data, stand volume data, and stand height data as well as other methods. Good summaries of these methods can be found in Clutter et al. (1983), Carmean (1975), and Hagglund (1981). Site index is by far the most common method in the United States for estimating site quality. Although the ability of this method to measure site quality has been disputed (Mader, 1963; Sammi, 1965; and others), it remains the best accepted method.

The first site index curves were constructed via graphical methods (Bruce, 1926). These curves were anamorphic, the members of a given family of H/A curves all have an identical shape and are proportional to each other. Following this, statistical methods (typically regression analysis) replaced the graphical methods. Bull (1931) introduced the use of polymorphic H/A curves, which allow curves for sites of differing quality to have different shapes. Since that time, polymorphic site index curves have been developed for loblolly pine (Trousdell et al., 1974; Amateis and Burkhart, 1985; and Devan and Burkhart, 1982), slash pine (Zarnoch and Feduccia, 1984), monterey pine (Bailey and Clutter, 1974), Engelmann spruce (Brickell, 1966), grand fir (Stage, 1963), Douglas-fir (King, 1966) and many other species. One of the more common approaches to developing polymorphic site index curves is known as the parameter prediction (or random coefficients) approach. Polymorphism is introduced by expressing one or more coefficients as a function of site index. This approach is demonstrated by Brickell (1966) and Trousdell et al. (1974). Indeed, the trend has been towards the development of increasingly complex and localized equations.

As an alternative to developing more models for smaller regions, the proposal of Turnbull (1978) may be pertinent. He proposes that a full forecasting system should include, among other things: (1) a basic yield model and (2) a feedback model. Munro (1984) agreed with this idea by considering the key characteristic of "third generation of forest growth modeling" to be the incorporation of a feedback loop. It seems that the idea of a feedback model could also be applied to site index estimation. A logical conclusion is that a simple anamorphic equation can be updated or localized with specific plot data; the result being as good or better than a more complex polymorphic model. An alternative manner by which this problem may be examined is to consider the initial equation as being developed from "prior" information. The feedback model is then used to incorporate additional sample data into the initial equation. This feedback idea has already been incorporated into several forest estimation problems.

The most common methodology from which these feedback models have been developed is the Bayes/ empirical Bayes methodology. One of the earliest works in forestry which used this idea of a prior and a sample data set was by Dane (1965). He applied both decision theory techniques and Bayesian statistics to determine the usefulness of employing additional information in forest inventories. Burk and Ek (1980, 1982) have extensively applied both empirical Bayes and James-Stein methodology to simultaneous estimates of non-overlapping strata in forest inventories. Ek and Issos (1976, 1978) also applied Bayesian theory to problems in forest inventory. They point out that forest inventory relies heavily on prior information in the design of surveys and it is a logical extension to include prior information in the actual estimation problem.

Bayesian methods have also been used to adjust the parameters of a regional diameter increment model so that they are more appropriate for a specific subregion (Gertner, 1984). The diameter increment model which was localized was the one found in STEMS (Belcher, 1981). In an earlier paper, Gertner (1983) indicates that a growth projection system using a Bayesian localization/updating procedure is the initial stage in developing a system which can "learn". The diameter growth submodel in Prognosis (Wykoff et al., 1982) has also been used as an example of how to calibrate a growth model using additional sample information (Stage, 1981). Green and Strawderman (1985) used empirical Bayes methods to develop individual tree volume equations.

A second method which may be used to incorporate prior information is filtering. Filtering is the estimate of the current state of a system contingent upon all prior information and samples. Dixon and Howitt (1979) used a Kalman Filter in a forest inventory system. Essentially a Kalman Filter estimator of a vector of parameters for some linear model ($\underline{\beta}$) is given by:

$$\hat{\underline{\beta}}_i = \hat{\underline{\beta}}_{i-1} + K(Y - X\hat{\underline{\beta}}_{i-1})$$

where:

$\hat{\beta}_i$ = estimate of β for plot i.

$\hat{\beta}_{i-1}$ = estimate of β for plot $i - 1$.

K = Kalman gain matrix which is a function of prior and sample data.

This relationship holds equally well if $\underline{\beta}$ is a random vector or if it is fixed (Diderrich, 1985). Diderrich (1985) shows that the updating step of a Kalman filter is equivalent to the Goldberger-Theil mixed estimator commonly used in econometrics (Theil, 1963).

Another possible method of incorporating prior information is to simply pool the prior information and the sample data. A method for doing this is given in Beck and Arnold (1977).

Chapter III

METHODS AND MATERIALS

General Approach and Objectives

As previously stated, the objective of this study was to develop and assess the properties of localized site index equations. In order to accomplish this objective, several specific steps were taken:

1. Development or selection of an unbiased mean H/A model from the prior information.
2. Estimation of a variety of adjustment models to be applied to this simple H/A curve.
3. Comparison of the different adjustment models.
4. Development or selection of a more complex polymorphic H/A curve to be used as a comparison model.

5. Comparison of the best adjusted model with the best available polymorphic equation.

Data

A large amount of data was available for use in this project. Three major data sets were used:

1. Cutover-site loblolly pine plantation data.
2. Old-field site loblolly pine plantation data.
3. Natural loblolly stand data.

Cutover-site Plantation Data

These data were collected in the 1981-1982 dormant season from cutover, site-prepared plantations in the Southeast. The data encompass a wide physiographic region. One dominant and one codominant tree were felled in each stand. These trees were free of damage and showed no visible signs of suppression. After felling, the trees were bucked into four-foot sections. The total height and age of each cross-section was then determined. Because of the inability to section trees such that the cross-section falls exactly at the end of a particular year's growth, a "bias" correction was used to adjust the height of each cross-section. The correction employed was one presented by Carmean (1972). One-half of the year's leader length was added to the height at the top of the appropriate section. The year's leader length was estimated as an average of the annual growth of two 4-foot bolts above and below the section. Dyer and Bailey (1987) found this method of correction to be best of six different methods examined. The cutover-site plantation data consisted of 4231 H/A pairs. Each pair of trees per plot was combined as one "average site tree" per plot for the purpose of estimation. There were 182 such plots (or stands) represented in the data. A summary of the data is presented in Table [1].

In addition to these stem analysis data, a permanent plot was also established in each of the 182 stands. Actual average height of the dominant and codominant trees were available for the initial measurement and the first (3-year) remeasurement. Summaries of the initial measurement and the first three year remeasurement data are also presented in Table [1].

Table 1. Summary of cutover-site loblolly pine plantation data.

Number of Observations	Variable	Minimum Value	Mean	Maximum Value	Standard Deviation
Stem Analysis Data					
182	AGE	8.00	15.10	25.00	4.15
182	AVGDBH	3.35	6.65	11.33	1.57
182	AVGHT	17.70	42.07	73.85	4.20
Permanent Plot Data Initial Measurement					
182	AGE	8.00	15.14	25.00	4.15
182	AVGQMD	2.79	5.86	9.54	1.23
182	AVGHD	15.00	40.53	74.00	11.88
Permanent Plot Data 3-Year Remeasurement					
177	AGE	11.00	18.23	28.00	4.09
177	AVGQMD	4.32	6.74	10.81	1.20
177	AVGHD	26.00	49.61	82.00	11.18

Where:

AGE = Age since planting (years)
AVGDBH = Average diameter at breast height (inches)
AVGHD = Average height of the dominant and codominant trees (feet)
AVGHT = Average individual tree height (feet)
AVGQMD = Average quadratic mean diameter (inches)

Old-field Plantation Data

These data were gathered in 1969-1970 and consist of stem analysis trees. One dominant and one codominant tree were felled in each stand. These trees were free of damage and showed no visible signs of suppression. As in the cutover-site plantation data, the trees were again sectioned at 4-foot intervals and Carmean's (1972) height correction technique was applied to each section. The two trees per plot were averaged to yield one "average site tree". A complete description of these data is given in Burkhart et al. (1972b). A summary is presented in Table [2]. Temporary plot data are also available, but no remeasurement data were taken. These data are also included in Table [2].

Natural Stand Data

These data were also gathered in 1969-1970 and consist of stem analysis and temporary plot data. The stem analysis trees were sectioned at 4-foot intervals and Carmean's (1972) height correction technique was applied to each section. A complete description of these data is given in Burkhart et al. (1972a). A summary is presented in Table [3].

Table 2. Summary of old-field loblolly pine plantation data

Number of Observations	Variable	Minimum Value	Mean	Maximum Value	Standard Deviation
Stem Analysis Data					
178	AGE	10.00	16.94	32.00	4.58
178	AVGHT	24.50	44.54	80.60	10.95
Temporary Plot Data					
240	AGE	8.00	15.54	35.00	4.84
240	AVGQMD	2.86	5.88	11.27	1.37
240	AVGHD	20.50	42.96	87.50	12.04

Where:

AGE = Age since planting (years)
AVGQMD = Average quadratic mean diameter (inches)
AVGHT = Average height of stem analysis trees at time of felling (feet)
AVGHD = Average height of dominant and codominant trees (feet)

Table 3. Summary of natural stand loblolly pine data.

Number of Observations	Variable	Minimum Value	Mean	Maximum Value	Standard Deviation
Stem Analysis Data					
56	AGE	13.5	25.67	41.00	7.14
56	AVGHT	39.05	57.17	87.50	9.75
Temporary Plot Data					
121	AGE	13.00	29.58	77.00	11.76
121	AVGQMD	4.51	7.88	15.30	1.98
121	AVGHD	39.50	61.00	90.00	10.39

Where:

AGE = Age since planting (years)
AVGQMD = Average quadratic mean diameter (inches)
AVGHT = Average height of stem analysis trees at time of felling (feet)
AVGHD = Average height of dominant and codominant trees (feet)

Methods of Analysis

Initial Model

One of the simplest and yet quite accurate models for describing the H/A development is the Schumacher type model:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad [2.0]$$

where:

$\underline{Y} = [\, \underline{\log(H)}]$

$X = [\underline{1}, \underline{A}^{-1}]$

$\underline{\beta} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ = parameter to be estimated

$\underline{\varepsilon}$ = vector of i.i.d. errors

H = height of individual tree, average stand height, or other appropriate height measurement

A = tree age, plantation age, or other appropriate age measurement

log = natural logarithm

Equation [2.0] was the general guide curve model to be localized. The parameter estimates can be obtained several ways. They can be determined by the average total height and age value of each plot. The parameters for this general model with only a portion of the 182 plots (referred to as the fitting data) and the remainder was reserved for application of the adjustment mechanism and validation. One problem with fitting in this manner is that if a site-age correlation exists, there will be a bias in the associated curves. Generally the correlation will be negative and the associated curves will be too flat. A second approach would be to fit equation [2.0] to all cross-sectional height age pairs (again using only the fitting data set). This approach, although possibly ameliorating the site-age correlation, may give unwarranted weight to the larger trees. Therefore it may be necessary to weight the observations in some manner. Finally, a third approach would be to use previously published parameter estimates which have been shown to yield unbiased site index curves. All three approaches were considered and evaluated in this research.

Adjustment Methods

Initially three adjustment models were examined:

1. Pooling of Prior and Sample Information.
2. Empirical Bayes Estimation.
3. Kalman Filter Estimation.

Pooling of Prior and Sample Information

Beck and Arnold (1977) present an estimator for the linear model:

$$\underline{Y} = \underline{X}\beta_r + \underline{\varepsilon} \qquad [3.0]$$

where :

$\beta_r = a_1$, a scalar parameter

$\underline{X} =$ n by 1 vector of regressor variables

$\underline{Y} =$ n by 1 vector of dependent variables

$\underline{\varepsilon} =$ n by 1 vector of independent, normally distributed errors with mean $\underline{0}$ and covariance $\sigma^2 I$

which incorporates prior information. The estimator is equivalent to a least squares estimator using the combined data (prior and sample data). Its advantage is purely computational. By taking advantage of the results of the analysis of the prior data, a reduced number of calculations can be used to estimate the parameters of the combined data. Although the procedure given in Beck and Arnold (1979) is designed for Equation [3.0], similar estimators can be developed for models having more than one parameter. For Equation [2.0], our initial H/A model, the estimators combining prior and sample data for a_0 and a_1 are, respectively:

$$a_{o,c} = \left[\frac{n_o\overline{Y}_o + n_s\overline{Y}_s}{n}\right] - a_{1,c}\left[\frac{n_o\overline{X}_o + n_s\overline{X}_s}{n}\right] \qquad [4.0]$$

$$a_{1,c} = \frac{[n_o - 1]a_{1,o}S^2(X_o) + [n_s - 1]a_{1,s}S^2(X_s) + n_o\overline{X}_{o,*}\overline{Y}_{o,*} + n_s\overline{X}_{s,*}\overline{Y}_{s,*}}{[n_o - 1]S^2(X_o) + [n_s - 1]S^2(X_s) + n_o\overline{X}^2_{o,*} + n_s\overline{X}^2_{s,*}} \qquad [5.0]$$

where:

n_o = number of observations in prior information data set.

n_s = number of observations in sample data set.

$n = n_o + n_s$

$\overline{Y}_o$ = average value of Y for prior information data set.

$\overline{Y}_s$ = average value of Y for sample data set.

$\overline{Y}$ = average value of Y for sample and prior data combined.

$\overline{X}_o$ = average value of X for prior information data set.

$\overline{X}_s$ = average value of X for sample data set.

$S^2(X_o)$ = sample variance of X for prior information data.

$S^2(X_s)$ = sample variance of X for sample information data.

$\overline{X}_{o,*} = [\overline{X}_o - \overline{X}]$

$\overline{Y}_{o,*} = [\overline{Y}_o - \overline{Y}]$

$\overline{X}_{s,*} = [\overline{X}_s - \overline{X}]$

$\overline{Y}_{s,*} = [\overline{Y}_s - \overline{Y}]$

Empirical Bayes

Since there are some fundamental differences between Bayesian and non-Bayesian estimation, it may be worthwhile to give a concise definition of Bayesian estimation.

In non-Bayesian estimation, parameters are considered to be fixed and inferences about them are based on the properties of their estimators. In Bayesian inference, the parameters are considered to be random variables. Inferences about them are based on the conditional distribution of the parameters given some fixed data (Bickel and Doksum, 1977). This data and prior information about the parameters (the prior distribution) are incorporated via Bayes theorem:

$$f(\theta/Y) = \frac{f(Y/\theta)\pi(\theta)}{\int f(Y/\theta)\pi(\theta)\delta\theta}$$

where:

$\theta =$ unknown random variable

$Y =$ sample data

$f(\theta/Y) =$ posterior distribution of θ.

$f(Y/\theta) =$ distribution of Y given θ.

$\pi(\theta) =$ prior distribution of θ.

Estimates of θ are determined a number of different ways. One common method is to maximize the posterior distribution with respect to θ.

As Swindel (1972) mentions, problems with Bayesian methods often surface when they are applied: while Bayes theorem itself is indisputable, the application of the theorem may prove to be quite difficult. The prior distribution is seldom known and must be estimated, either subjectively or objectively. A subjective prior may be as simple as the best guess of an expert (i.e. the prior distribution of θ is "thought" to be $N(\mu, \sigma^2)$).

The use of a subjective prior is quite controversial and may yield erroneous results (Bross, 1969). A better accepted method of specifying the prior distribution is to use a sample of data to empirically estimate the distribution. A good introduction to empirical Bayes analysis is given by Casella (1985).

Several different estimators can be derived via the Empirical Bayes (EB) approach. The differences between estimators are due to the alternate distributions which may be assumed for the parameters, whether the parameters are independent, and the function to be optimized when estimating the parameters. The loss function to be examined here will be the minimization of squared error. The first EB model is based upon the assumption that the parameters are normally distributed. Clemmer and Krutchkoff (1968) present estimators assuming a_0 and a_1 are distributed $N(\mu, \lambda^2)$ and $N(\nu, \eta^2)$, respectively, that a_0 and a_1 are independent, and that $\underline{\varepsilon}$ is i.i.d. These estimators are :

$$a_{0,EB1,i} = \frac{n_s \hat{a}_{0,i} \lambda^2 + \mu\sigma^2}{n_s \lambda^2 + \sigma^2} \qquad [6.0]$$

where:

μ = mean of a_0 for prior information data set.

λ^2 = variance of a_0 for prior information data set.

σ^2 = variance of $\underline{\varepsilon}$ for prior information data set.

$a_{0,EB1,i}$ = empirical Bayes estimate for plot i.

$a_{0,i}$ = OLS estimate of a_i from plot i.

Now, since λ^2, σ^2 ,and μ are not known, it is suggested that they be estimated by maximum likelihood:

$$\hat{\sigma}^2 = \frac{1.0}{n_0} \sum_{i=1}^{n_o} (Y_i - \hat{Y}_i)^2$$

$$\hat{\lambda}^2 = \hat{\sigma}^2 [\sum_{i=1}^{n_0} (a_{0,i} - \hat{\mu})^2]^{-1}$$

$$\hat{\mu} = \frac{1.0}{n_0} \sum_{i=1}^{n_0} a_{0,i}$$

The estimate for a_1 is

$$a_{1,EB1,i} = \frac{S_{xx} a_{1,i} \eta^2 + \nu\sigma^2}{S_{xx}\eta^2 + \sigma^2} \quad [7.0]$$

where:

$$S_{xx} = \frac{1.0}{n_s} \sum_{i=1}^{n_s} (X_i - \overline{X})^2$$

$a_{1,EB1,i}$ = empirical bayes estimate for plot i.

$a_{1,i}$ = Estimate from plot i.

ν = mean of $a_{1,EB1,i}$ for prior information data set.

η^2 = variance of $a_{1,EB1,i}$ for prior information data set.

Estimates for ν and η^2 are needed. The MLE's for ν and η^2 are:

$$\hat{\eta}^2 = \hat{\sigma}^2 [\sum_{i=1}^{n_o} (a_{1,i} - \hat{\nu})^2]^{-1}$$

$$\hat{\nu} = \frac{1.0}{n_o} \sum_{i=1}^{n_o} a_{1,i}$$

A second empirical Bayes approach most simply can be thought of as a way of estimating some random variable θ having unknown distribution $G(\theta)$. The estimator which minimizes $E(\hat{\theta} - \theta)^2$ is the mean of this posterior distribution or,

$$\tilde{\theta} = \frac{\int \theta P(\hat{\theta} \mid \theta) dG(\theta)}{\int P(\hat{\theta} \mid \theta) dG(\theta)} \qquad [8.0]$$

Empirical Bayes uses data from independent observations of θ to estimate $\tilde{\theta}$ without explicitly assessing $G(\theta)$. Krutchkoff (1972) shows that an estimator of $\tilde{\theta}$, $\hat{\theta}$, reduces to

$$\hat{\theta} = \sum_{i=1}^{n_o} W_i \hat{\theta}_i$$

where:

$$W_i = \frac{P(\hat{\theta}_* \mid \hat{\theta}_i)}{\sum_{i=1}^{n_o} P(\hat{\theta}_* \mid \hat{\theta}_i)}$$

$\hat{\theta}_* =$ estimate of θ in present case.

$\hat{\theta}_i =$ estimate of θ in case i.

This is essentially a weighting of $\hat{\theta}_i$ which is proportional to the difference between $\hat{\theta}_*$ and $\hat{\theta}_i$. Personal communication with the author led to the following estimator:

$$a_{j,EB2,*} = \frac{\sum_{i=1}^{n_o} W_i a_{j,i}}{\sum_{i=1}^{n_o} W_i} \qquad [9.0]$$

where:

$$W_i = \sum_{i=1}^{n_o} [\frac{1.0}{\delta_i} \{ \exp \frac{(a_{j,i} - a_{j,\bullet})^2}{2\delta_i^2} \}]$$

$a_{j,i}$ = estimate of a_j for i^{th} data group.

$a_{j,*}$ = estimate of a_j for data group of interest.

δ_i^2 = Estimate of the variance of the i^{th} data group.

δ_i^2 can be estimated as the ordinary least squares (OLS) estimate for each data group or the simplifying assumption that $\hat{\delta}_i^2 = \hat{\delta}_j^2$ for all i,j can be made and $\hat{\delta}_i^2$ can be estimated from the entire prior data set.

Yet a third approach using empirical Bayesian methodology would be to explicitly estimate the prior distribution. The estimators would then be those which maximized the likelihood function. Clemmer and Krutchkoff's (1968) paper again proves to be useful. They showed that:

$$E(\underline{\beta}/\underline{Y}) = E(\underline{\beta}/\hat{\underline{\beta}}) = \hat{\underline{\beta}} + \frac{\sigma^2}{S_{xx}} \frac{f'_g(\hat{\underline{\beta}})}{f_g(\hat{\underline{\beta}})}$$

where $G(\beta)$ is the unknown prior distribution of the vector of parameters, and $f_g(\hat{\underline{\beta}})$ is the marginal density of $\hat{\underline{\beta}}$ They present as a consistent estimator of this ratio $(\frac{f'_g(\underline{\beta})}{f_g(\hat{\underline{\beta}})})$ the following:

$$\frac{f'_n(\hat{\underline{\beta}})}{f_n(\hat{\underline{\beta}})} = \frac{\sum_{i=1}^{n_1} \{(\sin \frac{A_i}{A_i})^2 - (\sin \frac{B_i}{B_i})^2\}}{h \sum_{i=1}^{n_1} (\sin \frac{B_i}{B_i})^2}$$

where:

$$A_i = \frac{\hat{\underline{B}}_* - \hat{\underline{B}}_i - h}{2h}$$

$$B_i = \frac{\hat{\underline{B}}_* - \hat{\underline{B}}_i}{2h}$$

$$h = n_o(\frac{-1}{5}) \max[\{\frac{1}{n_o} \sum_{i=1}^{n} (\hat{\underline{B}}_i - \bar{\underline{B}}_*)\}^{\frac{1}{2}}, \frac{S}{\sqrt{S_{xx}}}]$$

$\underline{B}_*$ = Estimate for plot of interest.

$\underline{B}_i$ = Estimate for plot i.

$$S = \sqrt{\frac{1.0}{n_o - 2} \sum_{i=1}^{n_s} (Y_i - \hat{Y}_i)^2}$$

S_{xx} = Defined previously.

The EB estimator of a_0 from equation [2.0] is then given as:

$$a_{0,EB3} = \hat{a}_0 + \hat{\sigma}^2 \frac{f'_n(\hat{a}_0)}{f_n(\hat{a}_0)} \qquad [10.0]$$

Similarly, a EB estimate of a_1 from equation [2.0] can be expressed as:

$$a_{1,EB3} = \hat{a}_1 + \frac{\hat{\sigma}^2}{S_{xx}} \frac{f'_n(\hat{a}_1)}{f_n(\hat{a}_1)} \qquad [11.0]$$

An assumption with this approach is that a_0 and a_1 are independent and uncorrelated with one another. However, Monte Carlo comparisons of these estimators with estimators that allowed for interdependence revealed there was no apparent difference (Clemmer and Krutchkoff, 1968).

Kalman Filter

Kalman Filter theory, most commonly used in the engineering field, is essentially a sequential implementation of the Goldberger-Theil (GT) mixed estimator (Theil, 1963). Didderich (1985) shows that the updating step in a Kalman Filter is simply the GT estimator. This estimator combines a priori information and sample data in linear models. Two equivalant methods of obtaining these estimators are presented by Didderich. Prior information is expressed by the equation:

$$\hat{\underline{\beta}}_p = \underline{\beta}_p + \underline{e}_p \qquad [12.0]$$

where:

$\underline{e}_p$ = Independent and identically distributed random vector with mean $\underline{0}$ and covariance matrix W_p.

$\underline{\beta}_p$ = A vector of coefficients for some linear model based on the prior information data set.

The sample information is expressed as:

$$Y_s = X\underline{\beta}_s + \underline{e}_s \qquad [13.0]$$

where:

$\underline{e}_s$ = Normally distributed random vector with mean $\underline{0}$ and covariance matrix W_s.

$\underline{\beta}_s$ = A vector of coefficients for some linear model based on the sample data set.

The additional assumption that $\underline{e}_p$ $\underline{e}_s$ are uncorrelated is also placed on this system. These two methods are:

1. minimize:

$$(\underline{\beta} - \hat{\underline{\beta}}_p)'W_p^{-1}(\underline{\beta} - \hat{\underline{\beta}}_p) + (Y_s - X\underline{\beta})'W_s^{-1}(Y_s - X\underline{\beta})$$

with respect to β.

2. express [12.0] and [13.0] as:

$$[\frac{Y_s}{\beta_p}] = [\frac{X_s}{I}]\underline{\beta} + [\frac{e_s}{e_p}] \quad [14.0]$$

$$\underline{Y}_c = X_c\underline{\beta} + \underline{e}_c$$

and solve the equation using generalized least squares. The estimator obtained from either method is:

$$\hat{\underline{\beta}}_{KF} = \hat{\underline{B}}_p + K(Y_s - X_s\hat{\underline{\beta}}_p) \quad [15.0]$$

where:

$$K = W_pX'_s[W_s + X_sW_pX'_s]^{-1}$$

Yet another way of expressing $\hat{\underline{\beta}}_{KF}$ is:

$$\hat{\underline{\beta}}_{KF} = (X'_s W_s^{-1} X_s + W_p^{-1})^{-1}(X'_s W_s^{-1} \underline{Y}_s + W_p^{-1} \underline{\beta}_p)$$
$$= \theta_s \hat{\underline{\beta}}_s + \theta_p \hat{\underline{\beta}}_p$$

where:

$$\theta_s = (X'_s W_s^{-1} X_s + W_p)^{-1}(X'_s W_s^{-1} X_s)$$
$$\theta_p = (X'_s W_s^{-1} X_s + W_p)^{-1} W_p^{-1}$$

$\theta_s$andθ_p are proportional to the inverse of the variance of $\underline{\beta}_s$and$\underline{\beta}_p$, respectively. It can be shown that these methods are equivalent and are the best linear unbiased estimators (BLUE) for equation [14.0].

The Kalman filter estimator is simple and intuitive. Expressed as equation [15.0], the estimate is equal to the a priori estimate plus the difference between the actual and predicted heights for the sample data, multiplied by the Kalman gain matrix. This gain matrix is a weighting of the a priori covariance matrix and the covariance matrix of the sample data. It is easily seen that this estimator can be expressed in a sequential fashion.

Development of Comparison Models

In order to assess the performance of the adjusted model, a comparison model must be developed. Several flexible polymorphic models were developed and examined. The models examined were:

$$\log(H) = \log(S)[\frac{A_I}{A}]^{g_1} e^{g_2(A^{-1} - A_I^{-1})} \qquad [16.0]$$

$$\log(H) = h_0 + h_i A^{-h_2} \qquad [17.0]$$

$$H = j_1(1.0 - e^{-j_2 A})^{(\frac{1.0}{1.0 - j_3})} \qquad [18.0]$$

$$H = k_1(e^{\{-[\frac{k_2}{A^{k_3}}]\}}) \qquad [19.0]$$

Equation [16.0] was obtained from Amateis and Burkhart (1985) and was originally fitted in the difference form:

$$\frac{d\log(H)}{d\frac{1.0}{A}} = g_1 \frac{\log(H)}{\frac{1.0}{A}} + g_2 \log(H)$$

Equations [17.0], [18.0], and [19.0] were obtained from Bailey and Clutter (1974), Richards (1959), and Stage (1963), respectively.

In order that Equations [16.0] through [19.0] be polymorphic, the parameters of each equation were expressed as functions of site index.

The best, as judged by mean error, mean squared error, mean absolute error, and other pertinent criteria, will then be used as the "yardstick" by which the adjusted simple models will be compared.

Comparison Criteria

As stated previously, a "fitting" subset of the cutover-site data was be used to develop and estimate parameters for the adjustment and comparison models. After this stage, the models were compared using the validation portion of the data. At some future point, additional old-field plantation and natural stand data were used as additional validation data. The remeasurements on permanent cutover-site plantation plots were also used for validation purposes. The first remeasurement is currently available and the second remeasurement is being obtained in the 1986-1987 field season. The criteria by which the various models were compared are fairly straightforward and can be classified into two major groups:

1. The accuracy and precision of the model in estimating site index of a plot (as represented by the "average site tree").

2. The accuracy and precision of the model in estimating the height growth pattern of a plot across time.

Chapter IV

RESULTS

Development of Polymorphic Models

Two basic approaches were tried in developing a comparison polymorphic model. The first is the parameter prediction approach. The second is a difference equation approach.

The first approach required that each plot have a site index value so that parameters could be made site specific. Since the data from cutover-site plantations consist of young trees, it was not possible to obtain actual estimates of site index. In order to estimate site index, several functions of age were fitted to each plot. These all behaved very badly, in that they leveled off quite prematurely at unreasonable asymptotes. At this point, four, more flexible, nonlinear functions were fitted to each plot. These were

$$\text{Logistic, } H_i = \beta_1[1.0 + \beta_2 \exp\{\beta_3 A_i\}]^{-1} \qquad [20.0]$$

$$\text{Cilliers Van Wyck, } H_i = \beta_1[1 - \exp\{-\beta_2(A_i - \beta_3)\}] \qquad [21.0]$$

$$\text{Weibull, } H_i = \beta_1 \exp[-\beta_2/(A_i)^{\beta_3}] \qquad [22.0]$$

$$\text{Chapman Richard, } H_i = \beta_1[1 - \exp\{-\beta_2 A_i\}]^{\beta_3} \qquad [23.0]$$

The asymptote terms in each of these models, β_1 had to be fixed in order to obtain convergence. According to Brewer et al. (1985), setting this term is often preferable to allowing it to be determined by the data, even if the data can support it. Of the three values examined (100, 150, and 200 feet), 150 feet was chosen. A comparison of these models resulted in the selection of equation [23.0], the Chapman Richards model. See Table [4] for summary statistics.

Equation [23.0], fitted to each plot, was then used to estimate site index (base age 25). The second step in obtaining a polymorphic site index equation was to fit equations [20.0] - [23.0] to the entire data set. Polymorphism was induced by replacing β_1, β_2, and β_3 with functions of the estimated site index. A number of functions were tried, with the following three being most promising:

1). $\hat{\beta}_2 = \alpha_0 + \alpha_1 \hat{S}^2$

2). $\hat{\beta}_2 = \alpha_0 + \alpha_1 \hat{S} + \alpha_2 \hat{S}^2$

3). $\hat{\beta}_3 = \alpha_0 + \alpha_1 \hat{S}$

After initially examining equations [20.0] to [23.0], the analysis was limited to equation [23.0] for several reasons. All four models displayed similar relationships. Because equation [23.0] was consistent with the estimation of individual plot site index, further

Table 4. Summary statistics for equations [20.0], [21.0], [22.0], and [23.0] fitted to cutover-site data set.

Equation	Statistic	Mean Residual	Mean (Residual)	Mean |Residual|
20	mean	-0.400	16.11	3.25
	max	- .140	54.13	6.36
	min	- .720	1.75	1.07
21	mean	.002	6.34	1.97
	max	.032	36.56	5.30
	min	- .011	0.47	0.59
22	mean	.180	5.78	1.87
	max	.730	33.89	4.90
	min	- .050	0.29	0.37
23	mean	- .070	5.67	1.83
	max	.240	34.27	5.12
	min	- .230	0.47	0.39

consideration of equations [20.0]-[22.0] was abandoned. Summary results for equation [23.0] fitted with the aforementioned functions of site index are in Table [5]. To incorporate the first function of site index, the following model was fitted:

$$H_i = \beta_1[1 - \exp\{-(\alpha_1 + \alpha_2\hat{S})A_i\}]^{\beta_3}$$

One further refinement was also attempted. Equation [23.0] can be constrained, by solving [23.0] for β_1, so that at index age the predicted height equals site index.

$$\beta_1 = H_i[1 - \exp(-\beta_2 A_i)]^{-\beta_3}$$

At index age, this becomes

$$\tilde{\beta}_1 = S[1 - \exp(-\beta_2 A_I)]^{-\beta_3}$$

Replacing β_1 in equation [23.0] with $\tilde{\beta}_1$ yields the following,

$$H_i = S[\frac{1 - \exp\{-\beta_2 A_i\}}{1 - \exp\{-\beta_2 A_I\}}]^{\beta_3} \qquad [24]$$

This equation was also fitted to the cutover-site data as it is shown and with the aforementioned functions of $\hat{S}$ incorporated. These results are also presented in Table [5].

Two approaches were attempted in developing a polymorphic equation. The second approach was based on equation [16.0], originally developed by Amateis and Burkhart (1985),

$$\log(H_i) = \log(S)(A_I | A_i)^{g_1} \exp\{g_2(A_i^{-1} - A_I^{-1})\}$$

This model has the advantage that it can be fitted with OLS if it is expressed in its differential form,

Table 5. Summary of comparison of site index models fitted to cutover-site data set.

Equation	MSE [a]	Mean Residual [b]	Mean $(Residual)^2$	Mean $\|Residual\|$
23.0	30.0998	-.3200	27.3500	3.8500
24.0	14.4278	.0400	13.8800	2.9000
23.1 [c]	14.7343	.0103	14.1409	2.9365
23.2	14.7326	.0080	14.1553	2.9332
23.3	no convergence			
24.1	14.3675	-.0271	13.7142	2.8831
24.2	14.3498	-.0392	13.7027	2.8775
24.3	14.3117	-.05496	13.6139	2.8716
25.0	22.0549	3.2407	13.4065	3.2500

a/ MSE is on the scale the equation was fitted on. The other statistics are on a height scale.

b/ Residual = average residual for each plot.

c/ Equation 23.i is Equation 23.0 with the ith function of site index incorporated to induce polymorphism.

$$\frac{d\log(H_A)}{d(\frac{1}{A_A})} = g_1 \frac{\log(H_A)}{\frac{1}{A_A}} + g_2 \log(H_A) \qquad [25]$$

The dependent and independent variables were estimated by,

$$\frac{\log(H_{i+i}) - \log(H_i)}{1/A_{i+1} - 1/A_i}$$

and

$$H_A = \frac{(H_{i+1} + H_i)}{2}$$

$$A_A = \frac{(A_{i+1} + A_i)}{2}.$$

In addition to linearizing equation [16.0], this re-expression of the model is invariant to index age, and it does not require an estimate of site index in order to estimate g_1 and g_2. The results are summarized in Table [5].

Although it may appear at first that Equation [25.0] is worse than any others, actually it is only worse when examining the mean residual. Squared residuals and absolute residuals are quite comparable to the other models. Equations [23.0], [23.1], [23.2], and [23.3] each have several nonsignificant coefficients, as determined by asymptotic t-tests. Additionally, these models are not constrained so that height equals site index when age equals index age. Equations [24.2] and [24.3] also have nonsignificant coefficients so the choice can be reduced to choosing between Equations [24.0], [24.2] and [25.0]. These three models were graphed for site indices 40, 60, and 80, and very few differences were found. It would seem that, given the slight differences between these three models, equation [25.0] is preferred since it is constrained such that $H_i = S$ at $A_i = A_I$ and since it is base-age invariant. Therefore equation [25.0] was used as the "yardstick" by which the localized equations were judged. Parameter estimates for equation [25.0] fitted to the fitting portions of cutover-site plantation, old-field plantation, and natural stand pine data sets are presented in Table [6].

Localized Models

Prior Information

The first step in applying any of the "adjusted" models previously discussed is to obtain some estimate of model prior information. Depending upon the model, there are a variety of pieces of information about this distribution which must be obtained. It is beyond the scope of this research to explore the distributional form of the data, but it is assumed for several models that the prior distribution is normal. At minimum, there must be an estimate of the mean and variance of the coefficient vector β available. Some models require $X'X$ and $X'Y$ matrices of the prior data. Others require a number of estimates of the coefficient vector and its variance. Essentially, we can define two sources of this information.

The information, if it is a single estimate of β and its variance, can be obtained by fitting equation [2.0] to the temporary plot data. This is to be defined as the Plot-Level prior information. If multiple estimates of β and $V(\beta)$ are needed, these can also be obtained on a plot-level basis by estimating these from the cutover-site plantation, old-

Table 6. Parameter estimates, root mean square error and standard errors of parameter estimates for equation [25.0] fitted to cutover-site, natural, and old-field stem analysis data sets.

Data Set	RMSE	Estimate of (standard error of) g_1	g_2
Cutover-site	4.4456	-.1091 (.0068)	-2.0142 (.0767)
Natural	6.7535	-.0927 (.0106)	-3.1386 (.1807)
Old-field	5.0872	-.0701 (.0099)	-3.5007 (.1179)

field plantation, and natural stand data sets. As an alternative, these estimates can be obtained from the stem analysis data by fitting equation [2.0] to the stem analysis trees on each plot individually. Unfortunately, this poses a problem in that the stem analysis data contains, quite obviously, sections as young as one year. The influence of these data is quite extreme. Table [7] presents coefficient values estimate from a randomly selected plot and illustrates the effect of removing these data points. From trial and error, it was decided to eliminate sections of age less than eight years when fitting equation [2.0] to the stem analysis data on individual plots. Information obtained from stem analysis is referred to as stem-analysis prior information.

Regardless of whether the prior information is plot-level or stem-analysis level, it is obtained from a random sample of the overall data set. This sampling is performed so that a portion of the data set is reserved for applying the updating process and for validation of the models. The portion of the data which is reserved for updating and validation is referred to as the validation data.

Updating Models

The updating models can be classed into three groups based on the source of prior information and the source of the updating information.

Group 1: Plot-level Prior Information and Plot-level Updating Information

The pooling approach, equation [5.0], and the Kalman filter approach, equation [15.0], can both be applied with plot-level information only.

Table 7. Estimate for Plot 5104 of the coefficient vector of equation [2.0] obtained by deleting certain sections.

Deleting Sections less than age	a_1	a_2	Sample size
1	3.88	-3.99	34
2	4.14	-6.46	32
3	4.30	-8.29	31
4	4.56	-11.55	29
5	4.64	-12.61	27
6	4.65	-12.80	26
7	4.66	-12.95	25
8	4.69	-13.53	22
9	4.78	-15.03	19
10	4.82	-15.85	17
11	4.82	-15.92	16
12	4.83	-16.66	15
13	4.82	-15.84	13

Group 2: Plot-level Prior Information and Stem-Analysis Updating Information

The pooling, equation [5.0], the Kalman filter, equation [15.0], and empirical Bayes, equations [9.0] and [11.0] can be applied with this combination of information.

Group 3: Stem-analysis Prior Information and Stem-analysis Updating Information

The three empirical Bayes approaches can be fitted using this particular data structure. Equations [5.0], [15.0], [7.0], [9.0], and [11.0] were fitted according to these three groups.

The equations were referred to as they are designated in Table [8]. This table summarizes the equations and the data structures to which they were fitted.

All models were fitted to the cutover-site plantation,[1] old-field plantation, and natural stand data sets. Since height-age curves are typically scaled to a height at a given index age, the most appropriate way to compare these models was to scale them in a similar manner. Unfortunately, an observed site index (height at age 25) was not available. Two reasonable options seemed available: to scale each equation to a common age (oldest available age); to scale them to the oldest height/age observation available on the plot. This latter method was chosen rather than selecting an arbitrary common age. All models were updated either with plot-level data (models [K] and [L]) or stem analysis data (remaining models). They were then compared with the remeasurement plot-level data or the last six sections of stem analysis data per plot (these were not used for updating).

---

[1] Because of the need for remeasurement data, models [K] and [L] could only be fitted to the cutover-site plantation data.

Table 8. A summary of the equations fitted and the data structures involved.

Model	Equation	Prior Data	Updating Data
A	15.0	Plot-Level	Stem-Analysis
B	5.0	Plot-Level	Stem-Analysis
D	9.0	Plot-Level	Stem-Analysis
F	11.0	Plot-Level	Stem-Analysis
H	7.0	Stem-Analysis	Stem-Analysis
I	9.0	Stem-Analysis	Stem-Analysis
J	11.0	Stem-Analysis	Stem-Analysis
K	5.0	Plot-Level	Plot-Level
L	15.0	Plot-Level	Plot-Level
M	Schumacher equation based only upon prior data		
N	25.0	Comparison polymorphic model	

The specific statistics used to compare the models were,

1). Overall Plot-level Bias, $\frac{1}{n_p}\sum_{i=1}^{n_p}(\frac{H_i - \hat{H}_i}{H_i}) \times 100$

where:

H_i = height of the dominant and codominant portion of the stand at time i, the first remeasurement

n_p = number of plots.

2). Overall Stem-Analysis Bias, $\frac{1}{n_{stem}}\sum_{i=1}^{n_{stem}}(\frac{H_i - \hat{H}_i}{H_i}) \times 100$

where:

H_i = height of stem analysis section i.

n_{stem} = number of sections in complete data set.

3). Mean Bias Across Plots, $\frac{1}{n_p}\sum_{j=1}^{n_p}\left[\sum_{i=1}^{n_{stem,j}}(\frac{H_{ij} - \hat{H}_{ij}}{H_{ij}}) \times 100\right]$

where:

H_{ij} = height of ith section on jth plot.

$n_{stem,j}$ = number of sections on plot j.

n_p = number of plots.

These statistics are presented in Tables [9], [10], and [11] for the three groups of models listed previously fitted to the cutover-site data set. Similar tables for the old-field plantation and natural stand data sets can be found in the appendix (Tables [13] - [18]).

Table 9. Summary statistics for updating models - Group 1, cutover-site loblolly pine data set.

Model	Overall Plot-level bias (variance of)	Overall Stem-Analysis bias (variance of)	Mean Bias Across Plots (variance of)
K	4.9167 (38.5125)	24.9505 (1436.83)	27.2334 232.86
L	4.8968 (38.8533)	24.9316 (1437.52)	27.2436 233.62
M	4.8941 (38.9869)	24.9361 (1437.60)	27.2445 233.71
N	3.5947 (43.7643)	1.1407 (655.07)	1.4375 198.50

Table 10. Summary statistics for updating models - Group 2, cutover-site loblolly pine data set.

Model	Overall Plot-level bias (variance of)	Overall Stem-Analysis bias (variance of)	Mean Bias Across Plots (variance of)
A	4.4124 (21.42)	- .2387 (47.11)	.4221 (10.86)
B	4.6865 (19.84)	- .3947 (46.25)	.2586 (9.98)
D	6.4313 (31.43)	-1.6251 (62.44)	-1.2520 (15.58)
F	1.7254 (347.78)	.7023 (83.55)	.3015 (13.36)
M	4.5894 (20.84)	-.3109 (46.82)	.4435 (10.87)
N	1.4992 (24.60)	1.0306 (45.27)	1.2650 (8.93)

Table 11. Summary statistics for updating models - Group 3, cutover-site loblolly pine data set.

Model	Overall Plot-level bias (variance of)	Overall Stem-Analysis bias (variance of)	Mean Bias Across Plots (variance of)
H	4.2558 (65.44)	-.2581 (57.18)	-.5103 (11.05)
I	6.2563 (21.00)	-1.4499 (49.69)	-.8216 (11.10)
J	-4.6510 (2865.87)	3.0003 (191.19)	2.4545 (42.03)
M	4.5546 (20.79)	-.2151 (46.56)	.4435 (10.87)
N	1.4585 (24.57)	1.0306 (45.27)	1.2650 (8.93)

These tables of statistics contain enough information to narrow the examination process down somewhat. Essentially, the purpose of these statistics was to narrow the choice of models down to three, one from each group.

In Group 1, model [L] was selected for further analysis. Although model [K] was consistently slightly better with respect to mean and variance of percent bias, the differences were inconsequential. The functional form of model [L] is so logical and straightforward that it was selected for further analysis.

In Group 2, model [A] consistently does better with respect to mean percent bias. In a few instances, model [F] performed slightly better but the variance of percent bias for this model is always quite large. Therefore model [A] was selected for further analysis. Model [A], like model [L], is a Kalman filter updating equation, so consistency between groups 1 and 2 is maintained.

Model [H] was selected for additional analysis in Group 3 for several reasons. An examination of Tables [9], [10], and [11] shows that this model is best with respect to mean bias in most cases. Model [I] is slightly better with respect to variance, but the mean bias of model [I] is worse. Model [H] certainly has lower variance than does model [J].

Therefore models [A], [H], and [L] were scrutinized more closely. Mean and variance of residuals (expressed on a percentage basis) were calculated for each of these models for classes of average total age of stem analysis trees. These values are graphed in Figures [1] and [2] for the cutover-site data set. The same relationships are presented in Figures [14] - [17] in the Appendix for the old-field and natural pine data sets.

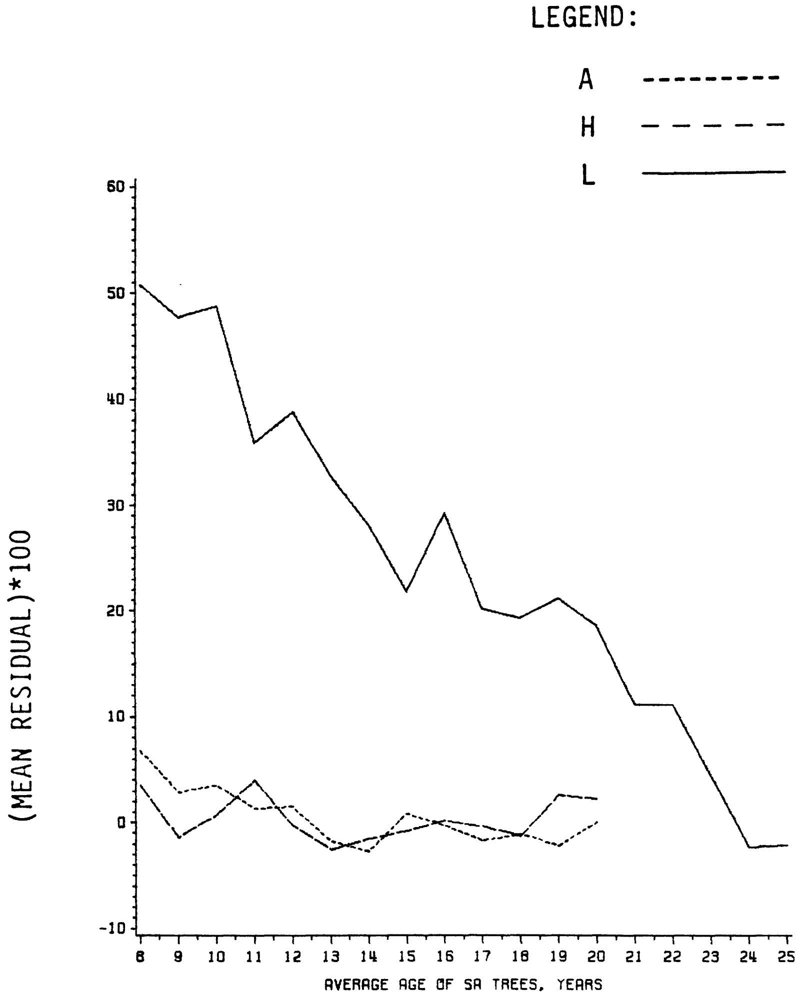


Figure 1. Mean percent residual versus average age of stem analysis trees used in updating for cutover-site loblolly pine data.

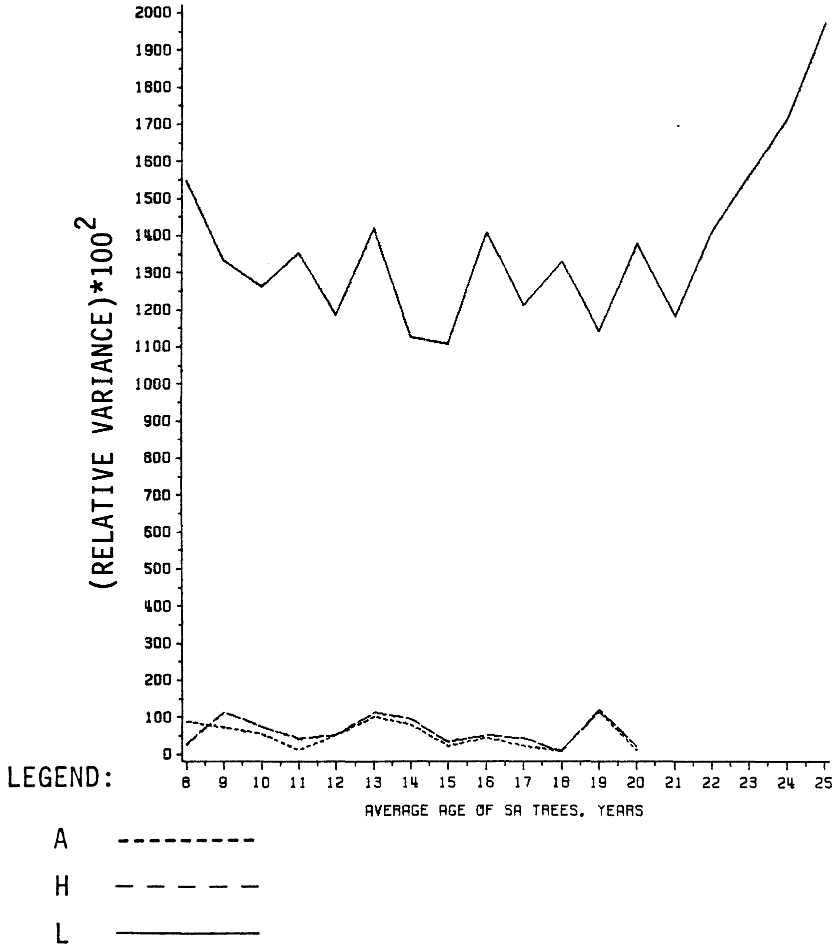


Figure 2. Variance of percent residuals versus average age of stem analysis trees used in updating for cutover-site loblolly pine

The most striking feature of these graphs is that model [L] is less satisfactory than the other two models. In the cutover-site plantation data set, there is a fairly clear trend across average age of the stem analysis trees used to compute the residuals, i.e., the older the tree, the more accurate the prediction. This relationship is not evident in the other two data sets. Unfortunately, such variables as the number of sections used to update the equation (applicable only to models [A] and [H]), the age of the initial section used to update the equation, and the age of the section upon which the residual is calculated mask many relationships. Figures [3] - [8] present these relationships for the cutover-site data set. Similar graphs are presented in the Appendix for the old-field and natural stand data sets (Figures [18] - [29]). In light of model [L]'s very poor performance in comparison with the other two models, further examination of model [L] was discontinued.

Figures [3] through [8] reveal that model [A] appears to have lower variance, whereas model [H] has a mean bias closer to zero, although it is subject to more extreme values. The bias of both models does not appear to be affected greatly by either the number of sections used in the updating step or the initial age at which updating is begun. However, there is a reduction in variance as both number of sections increases and initial age increases. There is, however, a trend in mean residuals across the age upon which the model is validated. In essence, the models predict poorer on younger sections than they do on older sections. The variance follows a similar although less pronounced trend.

The final step was to compare these two models with the unadjusted Schumacher equation (model [M]) and with the polymorphic equation (model [N]) developed by Amateis and Burkhart (1985). The overall bias and variance values for all updating equations are presented in Tables [9] through [11].

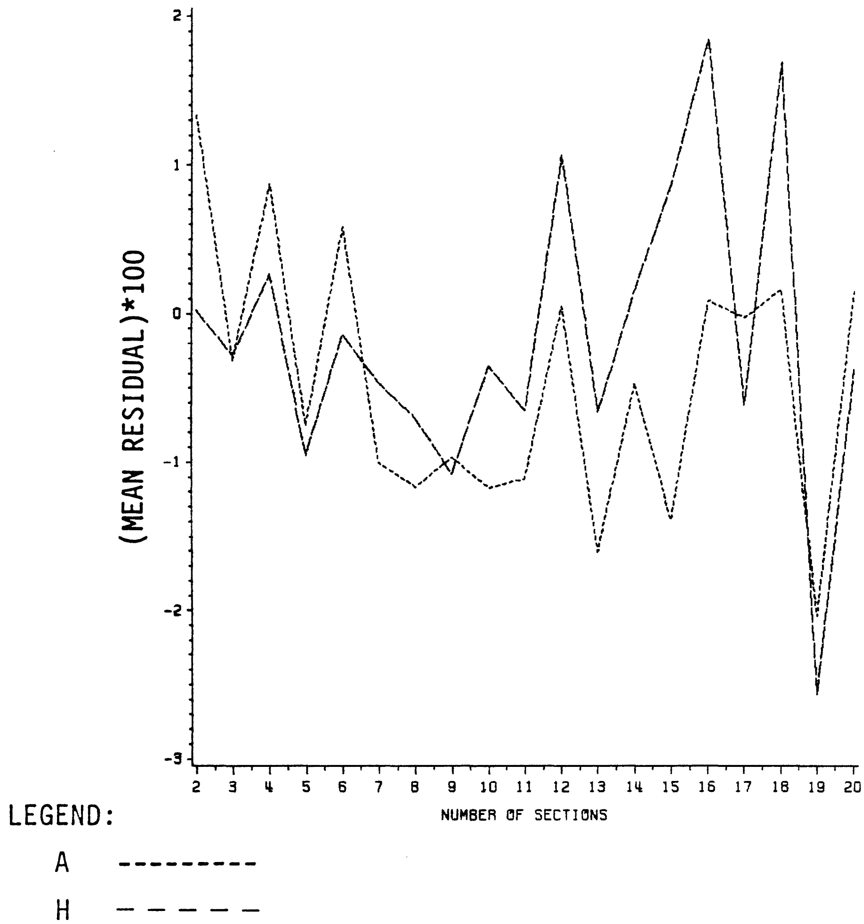


Figure 3. Mean percent residual versus number of sections used in updating for cutover-site loblolly pine data.

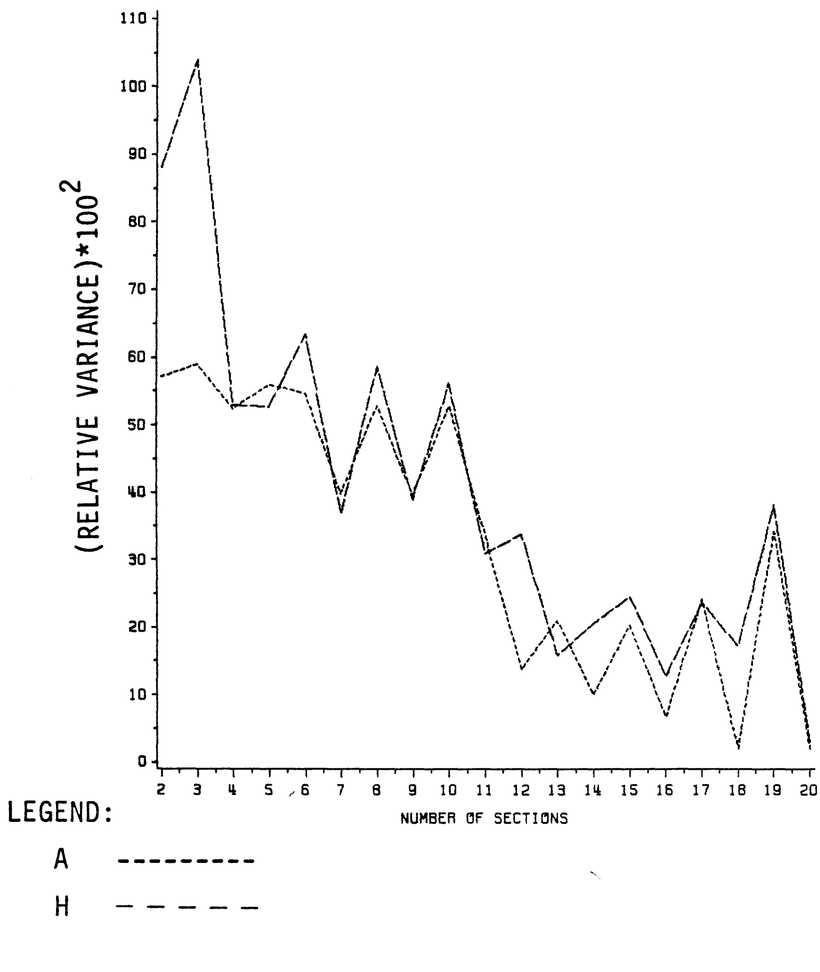


Figure 4. Variance of percent residuals versus number of sections used in updating for cutover-site loblolly pine data.

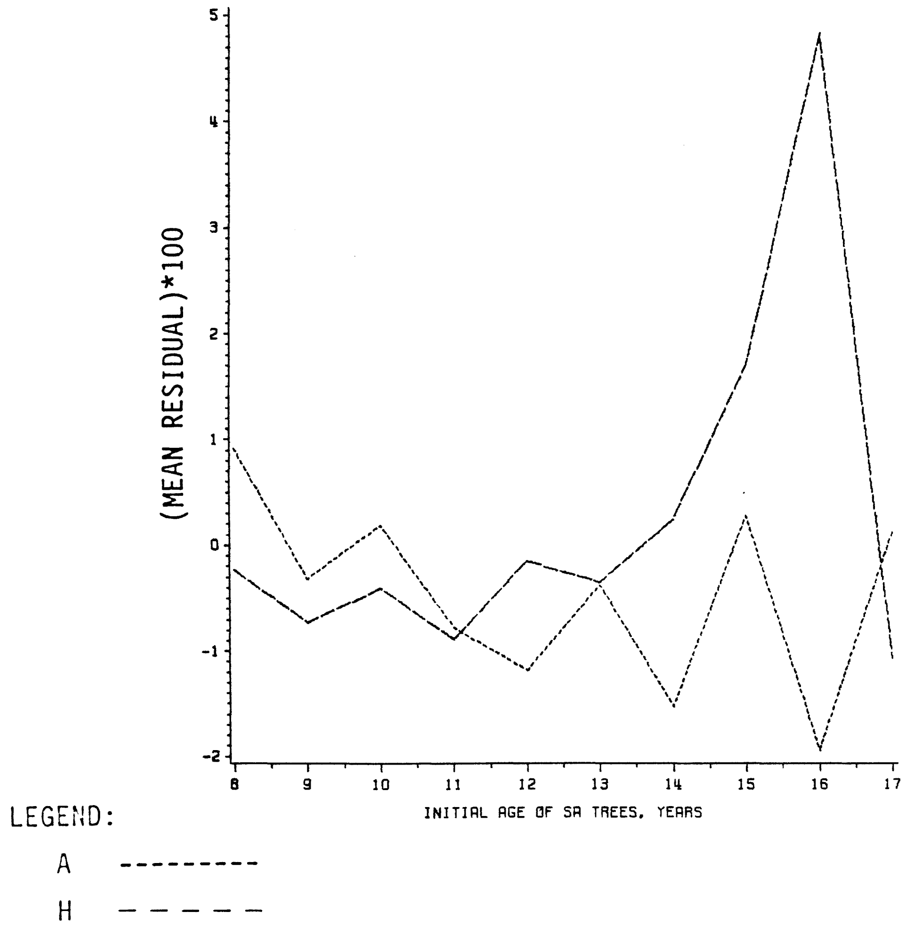


Figure 5. Mean percent residual versus initial age of stem analysis trees used in updating for cutover-site loblolly pine data.

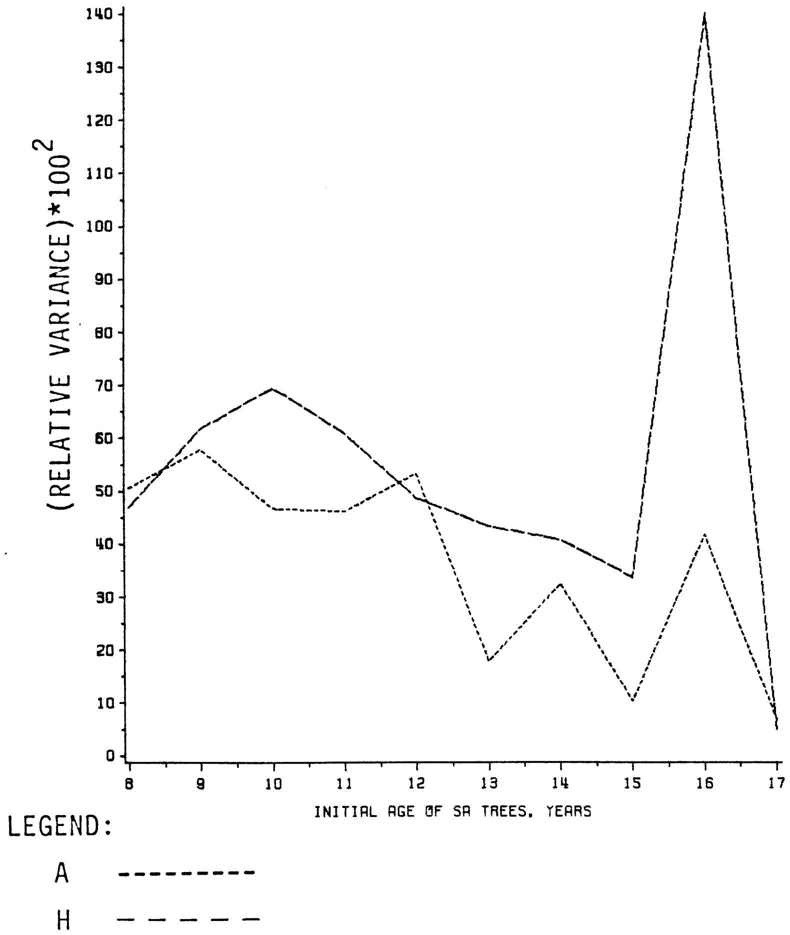


Figure 6. Variance of percent residuals versus initial age of stem analysis trees used in updating for cutover-site loblolly data.

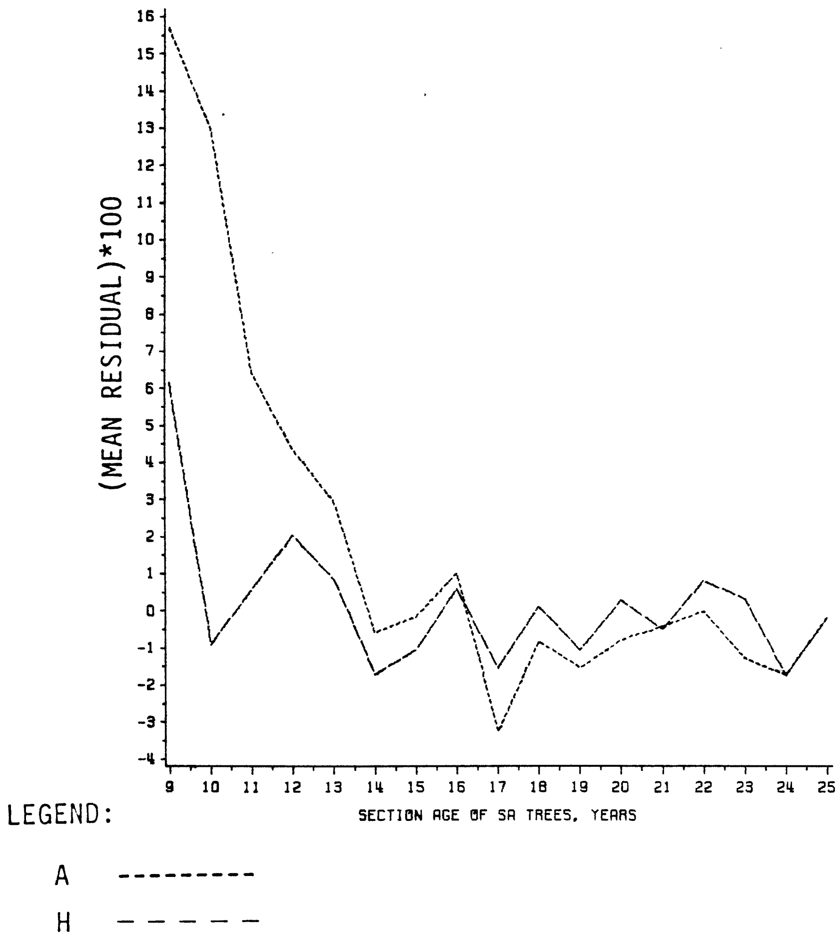


Figure 7. Mean percent residual versus age of sections used in updating for cutover-site loblolly pine data.

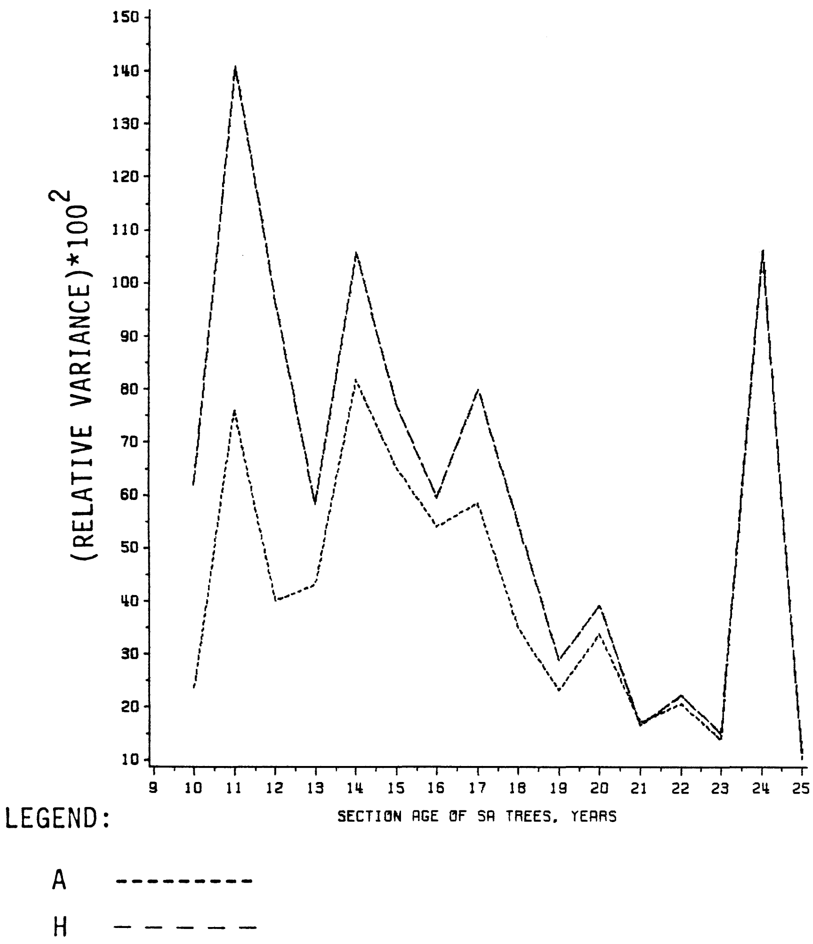


Figure 8. Variance of percent residuals versus age of section used in updating for cutover-site loblolly pine data.

At this point, it seems difficult to make a decision in favor of either model [A] or model [H] given the available material. However, there are reasons which would make the choice of model [A] preferable. The first reason is the source of prior information for the two models. Model [A] requires a single estimate of the coefficient vector and an estimate of the covariance matrix. Such information can be obtained from temporary plot data which is easily obtainable and is often available. Model [H], on the other hand, requires multiple estimates of the coefficient vector and covariance matrix. These can be obtained with multiple independent temporary plot data sets or through stem analysis, as done here. Neither of these are very desirable. In addition, the Kalman filter estimator is quite simple and intuitive. For these reasons, the predictive ability of model [A] was pursued in more detail.

Comparison of Model [A] with the Unadjusted Schumacher Model and with the Polymorphic Equation

In Tables [9] through [11] earlier in this text and Tables [13] through [18] in the appendix, the overall plot-level bias, overall stem-analysis bias, and mean stem-analysis bias across all plots were presented. Model [A] appears to be better than the unadjusted Schumacher equation (model [M]) according to all three statistics and across all three data sets. The polymorphic equation (model [N]) is better than model [A] for both the natural stand and old-field data sets, whereas it appears to be worse for the cutover-site data set.

Since the purpose of developing localized equations is to assist in the estimation of site index on a practical level, a more revealing way of examining model [A] was needed.

Therefore, two scenarios were considered. The first scenario was the case where the height of a tree is known at age 10. Four estimators were then compared in their ability to predict the height of the tree approximately 10 years later. The four estimators are:

1. Model [M]
2. Model [N]
3. Schumacher equation adjusted via a Kalman filter with previous height measurements available at ages 3 and 5
4. Schumacher equation adjusted via a Kalman filter with a previous height measurement available at age 5

The second scenario was the case where the the height of a tree is known at age 15. Five estimators were then compared as above. Again model [N] and Model [M] were examined, only they are of course constrained to go through age 15 rather than age 10. The other three estimators are

1. the Schumacher equation adjusted via a Kalman filter with previous height measurements available at ages 3, 5, and 10
2. the Schumacher equation adjusted via a Kalman filter with previous height measurements available at ages 5, and 10.
3. the Schumacher equation adjusted via a Kalman filter with a previous height measurement available at age 10.

The results of these two scenarios are presented in Table [12]. The most obvious relationship evident from this table is that having a height measurement available at age 15 greatly improves one's ability to predict heights at age 20. This is certainly intuitive and not very surprising. Secondly, the simple unadjusted Schumacher equation appears to perform better than the polymorphic comparison model. Thirdly, Kalman filter adjustment of the Schumacher equation is an improvement upon the unadjusted model as determined by most of the values in the table. As a graphical example of the information in this table, Figures 9-13 present site index curves for one randomly selected tree. Each graph presents site index curves based upon models [M], [N], and the Kalman filter model with specified amounts of prior information. Although one tree is certainly not adequate to prove the worth of the Kalman filter model, it is clear that the model is performing as expected. The curves are pulled towards the actual height pattern.

Table 12. A comparison of models [M], [N], and the Kalman filter estimator in their ability to predict total height of stem analysis trees, using the fitting portion of the cutover-site data as the prior information[1] and the validation portion to update and calculate residuals.

	Oldest Available Age = 10				Oldest Available Age = 15				
Model	Model [M]	[N]	Kalman filter adjusted with data at ages 3,5,10	5,10	Model [M]	Model [N]	Kalman filter adjusted with data at ages 3,5,10,15	5,10,15	10,15
mean residual	- 8.11	-14.15	- 5.64	10.43	- 1.70	-7.56	-1.29	0.91	0.92
mean |residual|	11.67	14.70	11.37	11.13	3.81	7.56	3.35	4.25	4.82
var (residual)	147.10	220.39	182.10	58.89	18.10	13.34	17.30	26.95	35.49

[1] Models [M] and [N] are based entirely on the prior data.

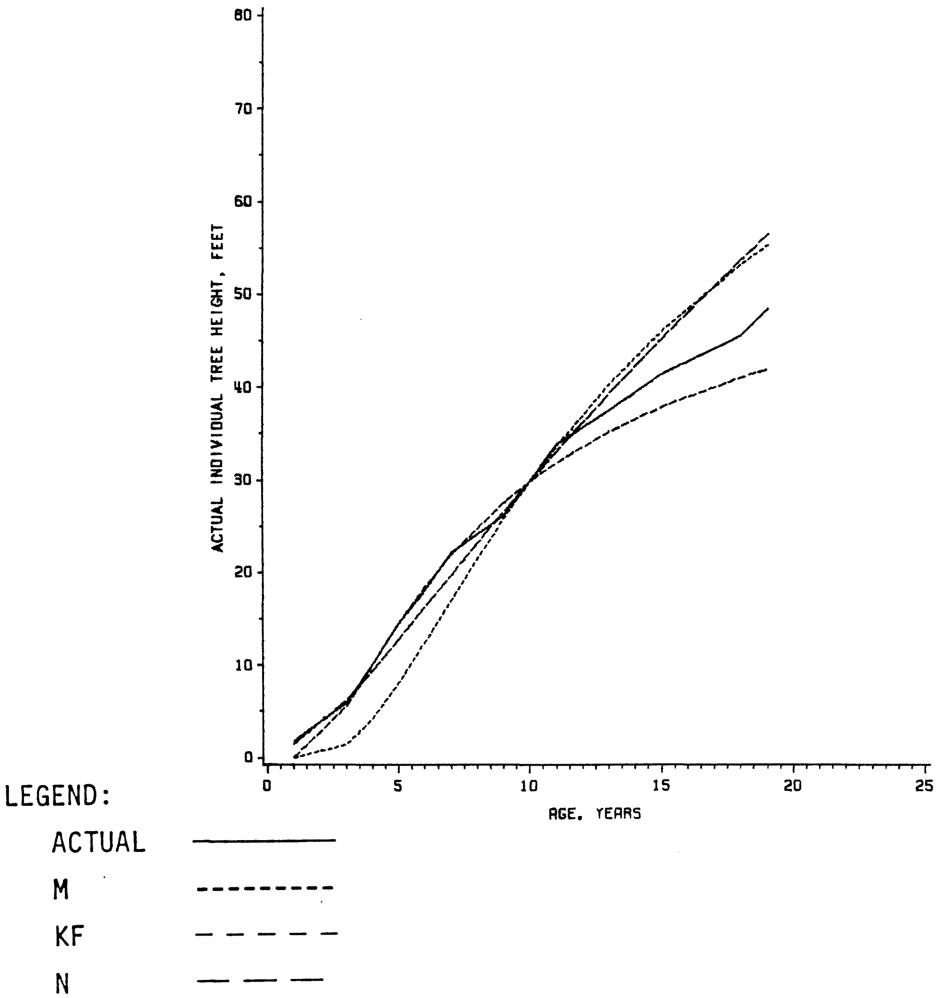


Figure 9. A comparison of site index curves for plot 1337, tree #1 with the heights at ages 5 and 10 used to update model [A].

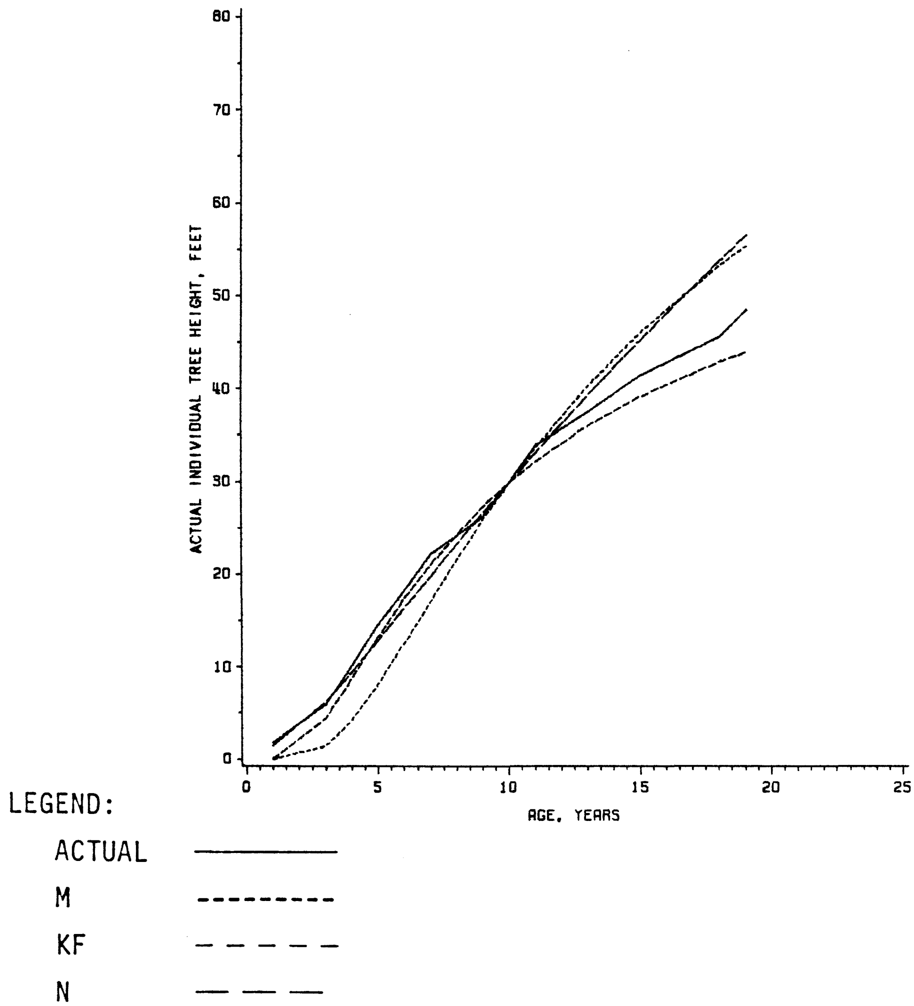


Figure 10. A comparison of site index curves for plot 1337, tree #1 with the heights at ages 3, 5, and 10 used to update model [A].

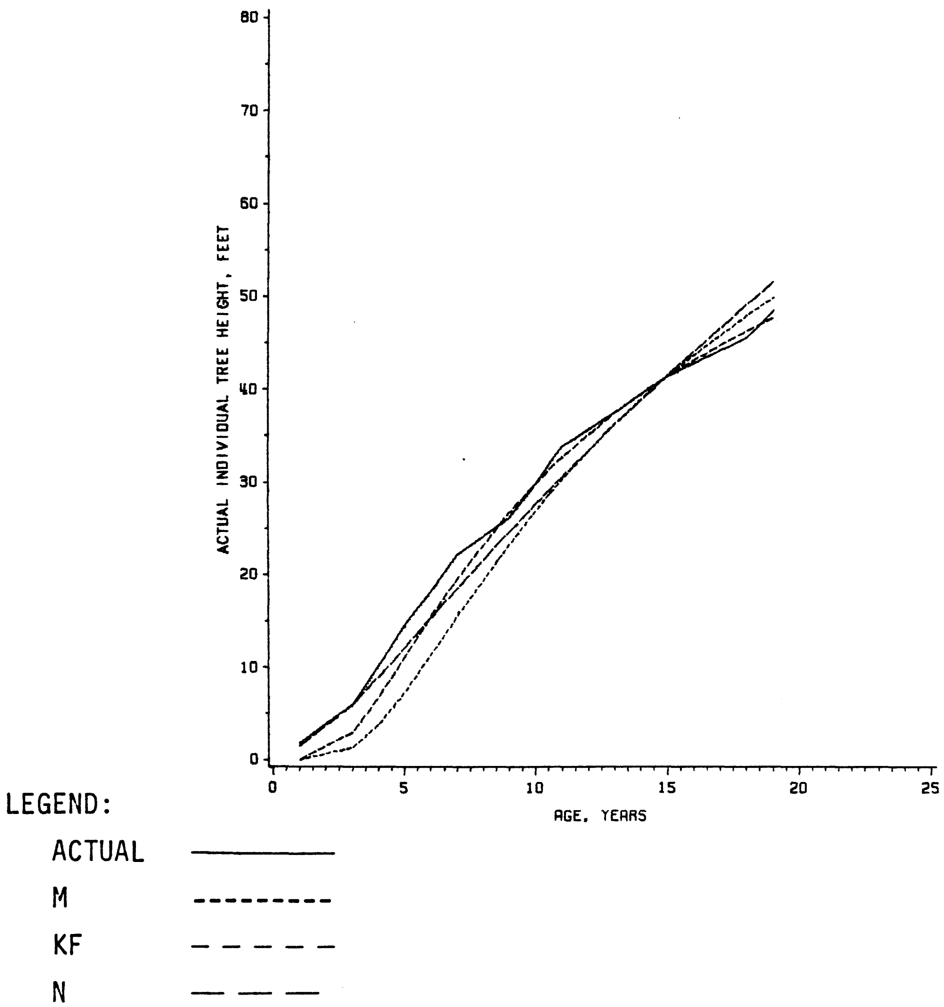


Figure 11. A comparison of site index curves for plot 1337, tree #1 with the heights at ages 10 and 15 used to update model [A].

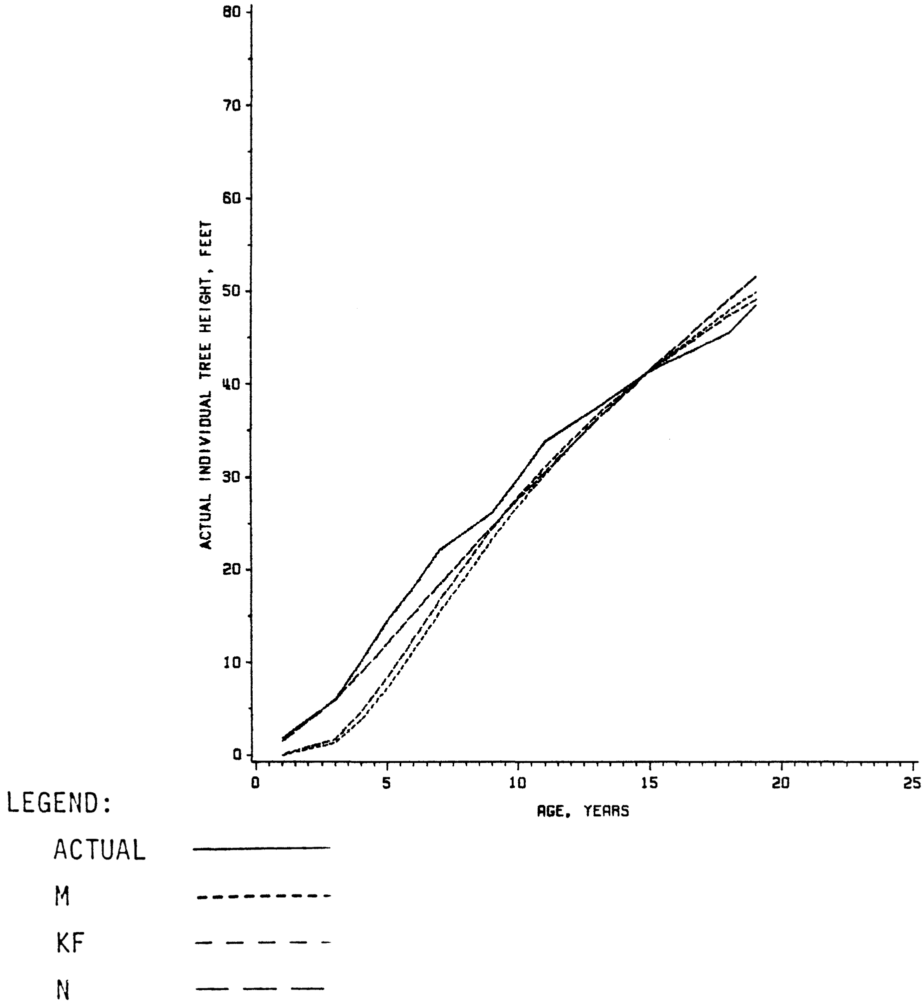


Figure 12. **A comparison of site index curves for plot 1337, tree #1 with the heights at ages 5, 10, and 15 used to update model [A].**

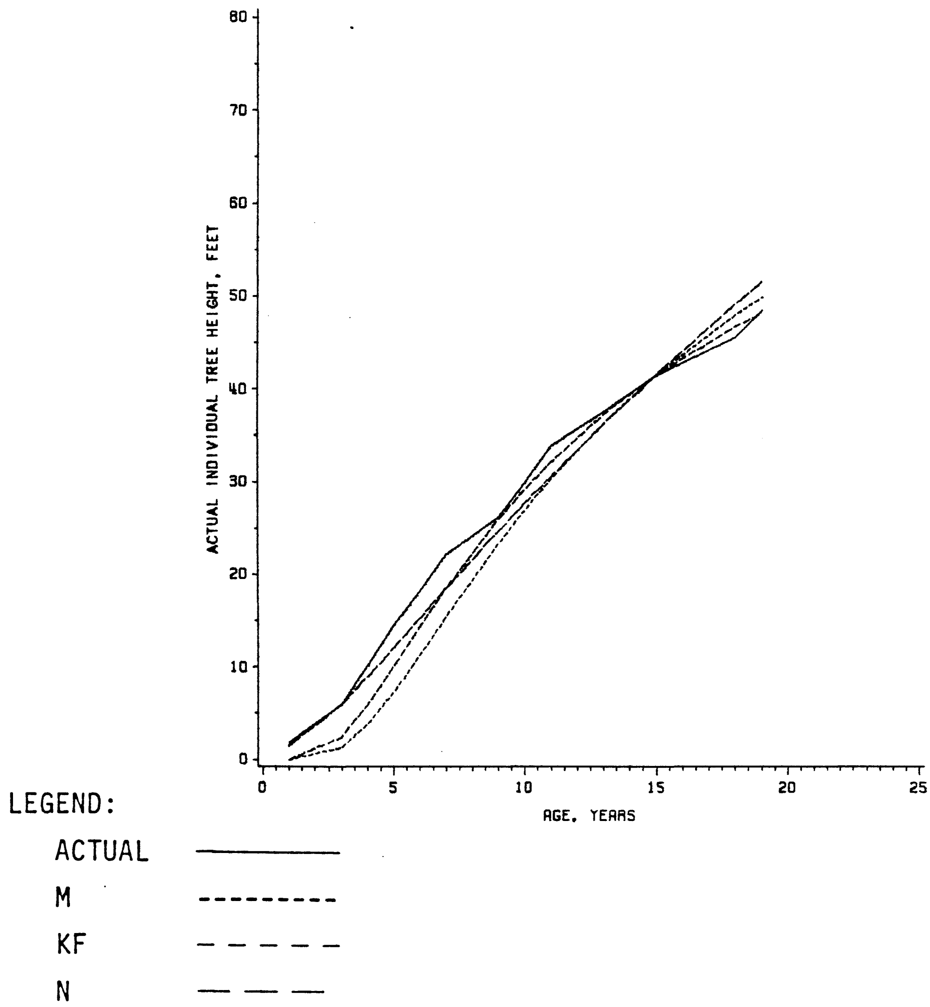


Figure 13. A comparison of site index curves for plot 1337, tree #1 with the heights at ages 3, 5, 10, and 15 used to update model [A].

Chapter V

SUMMARY AND CONCLUSIONS

A number of models for predicting the height-age development of forest stands were examined using three independent loblolly pine data sets. Several complicated models were examined and one selected as the base for comparison with the feedback models. The model which was chosen is a nonlinear, base-age invariant polymorphic equation. Feedback models based on empirical Bayes, pooling and Kalman filter theories were compared and contrasted. An empirical Bayes model and a Kalman filter model were selected on the basis of fit statistics such as average residual, average squared residual, and average absolute residual. These were then examined in more detail by plotting residuals for each model versus several parameters of interest. Few differences were observed and the Kalman filter model was selected as "best" on the basis of its simplicity and interpretability. This model is dependent upon having prior information about the height-age relationship in the form of an estimate of the coefficient vector and the variance of this estimate. The model then updates the equation by utilizing additional information. This information can be either stem analysis data or multiple plot-level measurements. In this study, stem analysis data was found to be most favorable. This was primarily due to only having two plot-level measurements. The Kalman filter model is an improvement upon the unadjusted height-age model and performs as well as a more complex polymorphic model.

Several possible applications exist for the feedback model presented here. If information on a large scale is unavailable for a specific area but information about similar areas is available, the height-age pattern in the specific area can be estimated via this method.

As specific examples of the use of this model, several scenarios were explored in which the height of a tree is known at a specific age, and there are various amounts of additional information available. The Kalman filter estimators compares favorably in these various scenarios.

Many models of forest growth are plagued by the problem of the predictions being inconsistent with newly obtained sample data. By incorporating these new data into the prediction process, this inconsistency is ameliorated to a certain extent. The height-age pattern of a stand, being vital to predictions of yield, is a logical place at which this incorporation should occur. The work presented here examines some methods of making this incorporation.

LITERATURE CITED

Amateis, Ralph L. and Harold E. Burkhart. 1985. Site index curves for loblolly pine plantations on cutover site-prepared lands. Southern Journal of Applied Forestry 9(3): 166-169.

Bailey , Robert L. and Jerome L. Clutter. 1974. Base-age invariant polymorphic site curves. Forest Science 20: 155-159.

Beck, James V. and Kenneth J. Arnold. 1977. Parameter estimation in engineering and science. John Wiley and Sons, N.Y. 501pp.

Belcher, David M. 1981. User's guide to STEMS (stand and tree evaluation and modeling system). U.S.D.A. Forest Service General Technical Report NC-70. 49 pp.

Bickel, Peter J. and Kjell A. Doksum. 1977. Mathematical statistics: Basic ideas and selected topics. Holden-Day, Inc., Oakland, Calif. 492pp.

Brewer, John A., Paul Y. Burns, and Quang V. Cao. 1985. Short-term projection accuracy of five asymptotic height-age curves for loblolly pine. Forest Science 31(2):414-418.

Bross, Irwin J. 1969. Applications of probability: Science vs. pseudoscience. Journal of the American Statistical Association 64:51-57.

Brickell, James E. 1966. Site index curves for engelmann spruce in the northern and central Rocky mountains. U.S.D.A. Forest Service Research Note INT-42. 8pp.

Bruce, Donald. 1926. A method of preparing timber-yield tables. Journal of Agricultural Research 32(6):543-557.

Bull, Henry. 1931. The use of polymorphic curves in determining site quality in young red pine plantations. Journal of Agricultural Research 43(1):1-28.

Burk, Thomas E. and Alan R. Ek. 1980. Application of James-Stein and empirical Bayes procedures to simultaneous estimation problems in forest inventory. Staff Paper Series Number 10, Dept. of Forest Resources, Univ. of Minnesota. 76pp.

Burk, Thomas E. and Alan R. Ek. 1982. Application of empirical Bayes/James-Stein procedures to simultaneous estimation problems in forest inventory. Forest Science 28(4):753-771.

Burkhart, Harold E., Robert C. Parker, and Richard G. Oderwald. 1972. Yields for natural stands of loblolly pine. VPI&SU Publ. FWS-2-72. 63pp.

Burkhart, Harold E., Robert C. Parker, Mike R. Strub, and Richard G. Oderwald. 1972. Yields of old-field loblolly pine plantations. VPI&SU Publ. FWS-3-72. 51pp.

Carmean, W.H. 1972. Site index curves for upland oaks in the Central states. Forest Science 18:102-120.

Carmean, Willard H. 1975. Forest site quality evaluation in the United States. Advances in Agronomy. 27:209-269.

Casella, George. 1985. An introduction to empirical Bayes data analysis. The American Statistician 39(2):83-87.

Clemmer, Bennie A. and Richard G. Krutchkoff. 1968. The use of empirical Bayes estimators in a linear regression model. Biometrika 55(3):525-534.

Clutter, Jerome L., James C. Fortson, Leon V. Pienaar, Graham H. Brister, and Robert L. Bailey. 1983. Timber management: A quantitative approach. John Wiley and Sons, N.Y. 333pp.

Clutter, Jerome L., and J. David Lenhart. 1968. Site index curves for old-field loblolly pine plantations in the Georgia Piedmont. Georgia Forest Research Council Report No. 22 - Series 1, November 1968. 4pp.

Dane, C.W. 1965. Statistical decision theory and its application to forest engineering. Journal of Forestry 63:276-279.

Devan, James S. and Harold E. Burkhart. 1982. Polymorphic site index equations for loblolly pine based on a segmented polynomial differential model. Forest Science 28(3):544-555.

Diderrich, George T. 1985. The Kalman filter from the perspectives of Goldberger-Theil estimators. The American Statistician 39(3):193-198.

Dixon, Bruce L. and Richard E. Howitt. 1979. Continuous forest inventory using a linear filter. Forest Science 25(4):675-689.

Dyer, Michael E. and Robert L. Bailey. 1987. A test of six methods for estimating true heights from stem analysis data. Forest Science 33(1):3-13.

Ek, Alan R. and J.N. Issos. 1976. Bayesian estimation methodology for forest inventory. Paper presented at Midwest Mensurationists Meeting, Sept 16-17, 1976, Glen Arbor, Michigan. 16pp.

Ek, Alan R. and James N. Issos. 1978. Bayesian theory and multi resource inventory. Integrated Inventories of Renewable Resources: Proceedings of the Workshop. USDA Forest Service General Technical Report RM-55. p. 291-298.

Gertner, George A. 1983. Application of a Bayesian technique for increasing the precision of growth projection estimates. IN: Forest Growth Modelling and Simulation, Vienna, Austria, p. 103-113.

Gertner, George Z. 1984. Localizing a diameter increment model with a sequential Bayesian procedure. Forest Science 30(4):851-864.

Green, Edwin J. and William E. Strawderman. 1985. The use of Bayes / empirical Bayes estimation in individual tree volume equation development. Forest Science 31(4):975-990.

Hagglund, Bjorn. 1981. Evaluation of forest site productivity. Commonwealth Forestry Bureau, Forest Abstracts Review Article 42(11):515-527.

King, James E. 1966. Site index curves for Douglas-fir in the Pacific Northwest. Weyerhauser Forestry Paper No. 8. 49pp.

Krutchkoff, Richard G. 1972. Empirical Bayes estimation. The American Statistician 26(5):14-16.

Madar, Donald L. 1963. Volume growth measurement: An analysis of function an characteristics in site evaluation. Journal of Forestry 61(3):193-198.

Munro, Donald D. 1984. Growth modelling for fast-growing plantations. IN: Proceedings IUFRO Symposium on Site and Productivity of Fast Growing Plantations, Pretoria and Pietermaritzburg, South Africa, Volume 1. p. 334-344.

Richards, F.J. 1959. A flexible growth function for empirical use. Journal of Experimental Botany 10(29):290-300.

Sammi, John C. 1965. An appeal for a better index of site. Journal of Forestry 67(3):174-176.

Schumacher, F.X. 1939. A new growth curve and its application to timber-yield studies. Journal of Forestry 37:819-820.

Stage, Albert R. 1963. A mathematical approach to polymorphic site index curves for Grand fir. Forest Science 9(2):167-180.

Stage, Albert R. 1981. Use of self-calibration procedures to adjust general regional yield. Paper presented at XVII IUFRO World Congress, S.4.01, Sept. 6-17, 1981. p. 365-375.

Swindel, Benee F. 1972. The Bayesian Controversy. USDA Forest Service Research Paper SE-95. 12pp.

Theil, H. 1963. On the use of incomplete prior information in regression analysis. Journal of the American Statistical Association 58(303):401-414.

Trousdell, Kenneth B., Donald E. Beck, and F. Thomas Lloyd. 1974. Site index for loblolly pine in the Atlantic Coastal Plain of the Carolinas and Virginia. USDA Forest Service Research Paper SE-115. 11pp.

Turnbull, Kenneth J. 1978. Long-term yield forecasting models: Validation and iterative estimation. IN: Growth Models for Long Term Forecasting of Timber Yields. FWS-1-78, School of Forest and Wildlife Resources, VPI and SU. p. 224-229.

Wykoff, William R., Nicholas L. Crookston and Albert R. Stage. 1982. User's guide to the stand Prognosis model. USDA Forest Service General Technical Report INT-133. 112pp.

Zarnoch, S. J., and D.P. Feduccia. 1984. Slash pine plantation site index curves for the West Gulf. Southern Journal of Applied Forestry 8(4):223-225.

APPENDIX

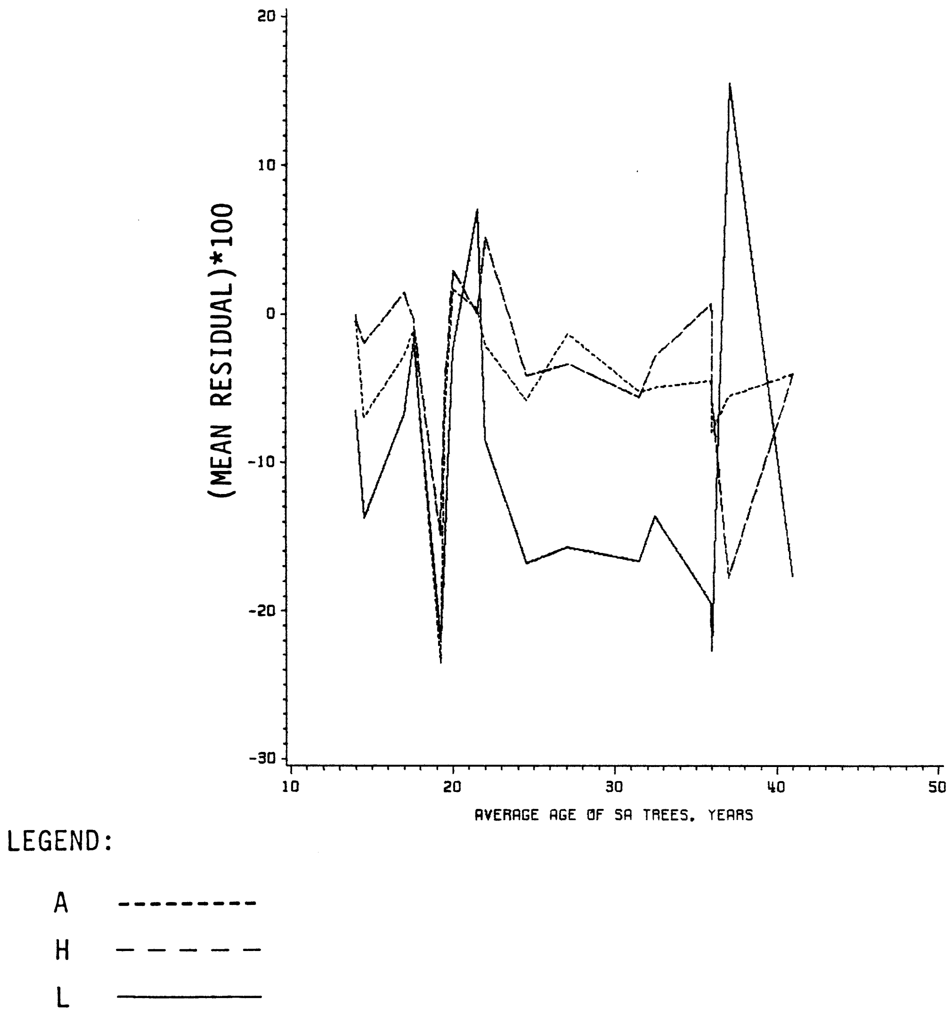


Figure 14. Mean percent residual versus average age of stem analysis trees used in updating for natural pine stand data set.

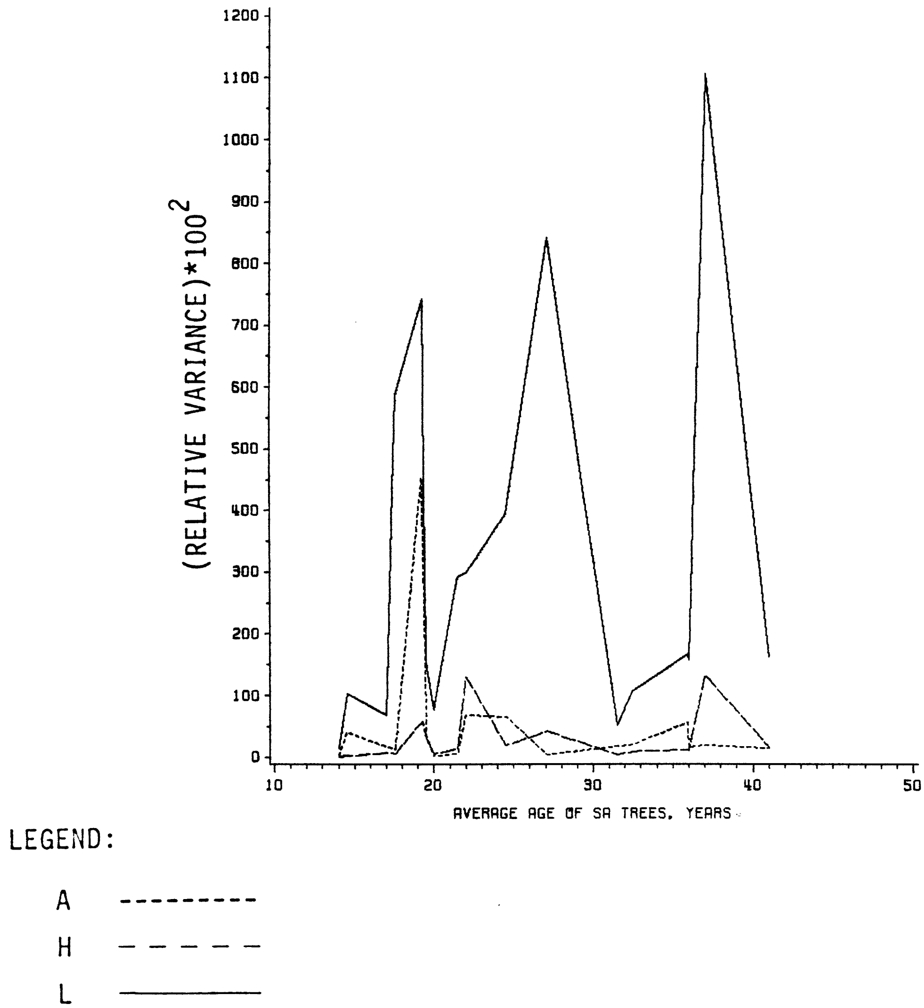


Figure 15. **Variance of percent residual versus average age of stem analysis trees used in updating for natural loblolly pine stand data**

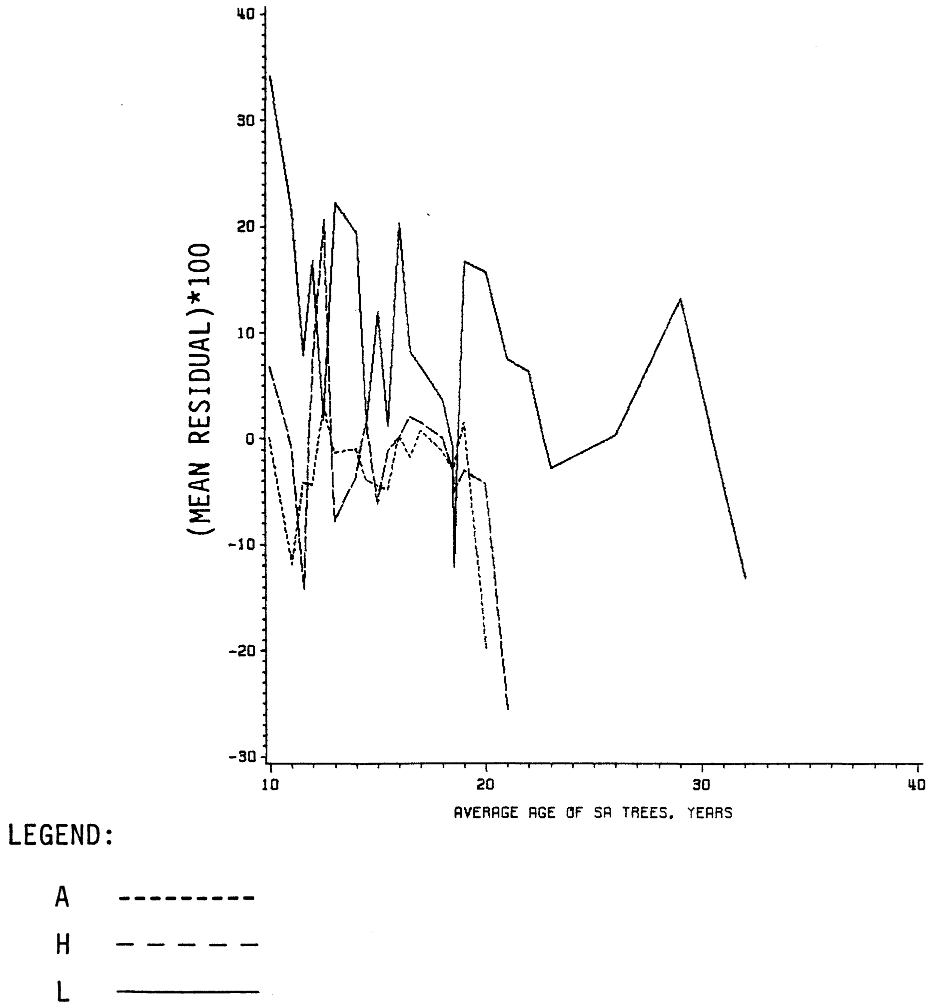


Figure 16. Mean percent residual versus average age of stem analysis trees used in updating for old-field plantation data set.

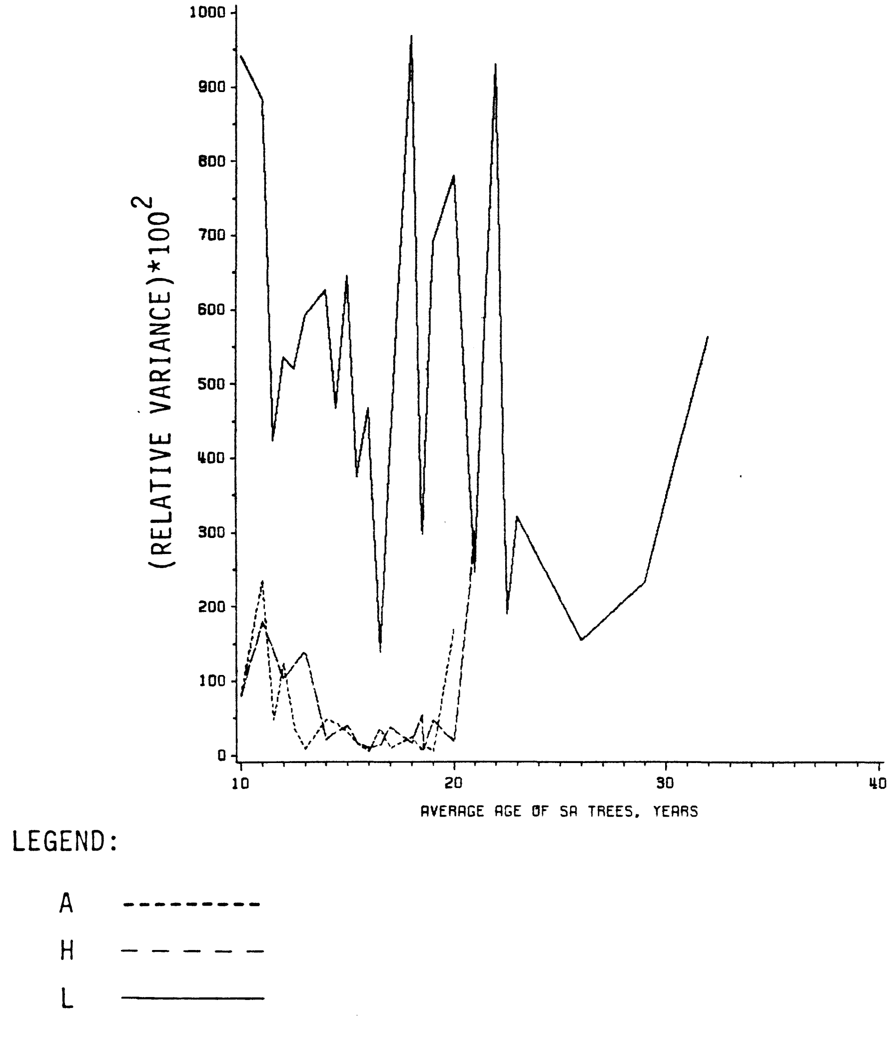


Figure 17. Variance of percent residual versus average age of stem analysis trees used in updating for old-field loblolly pine data set

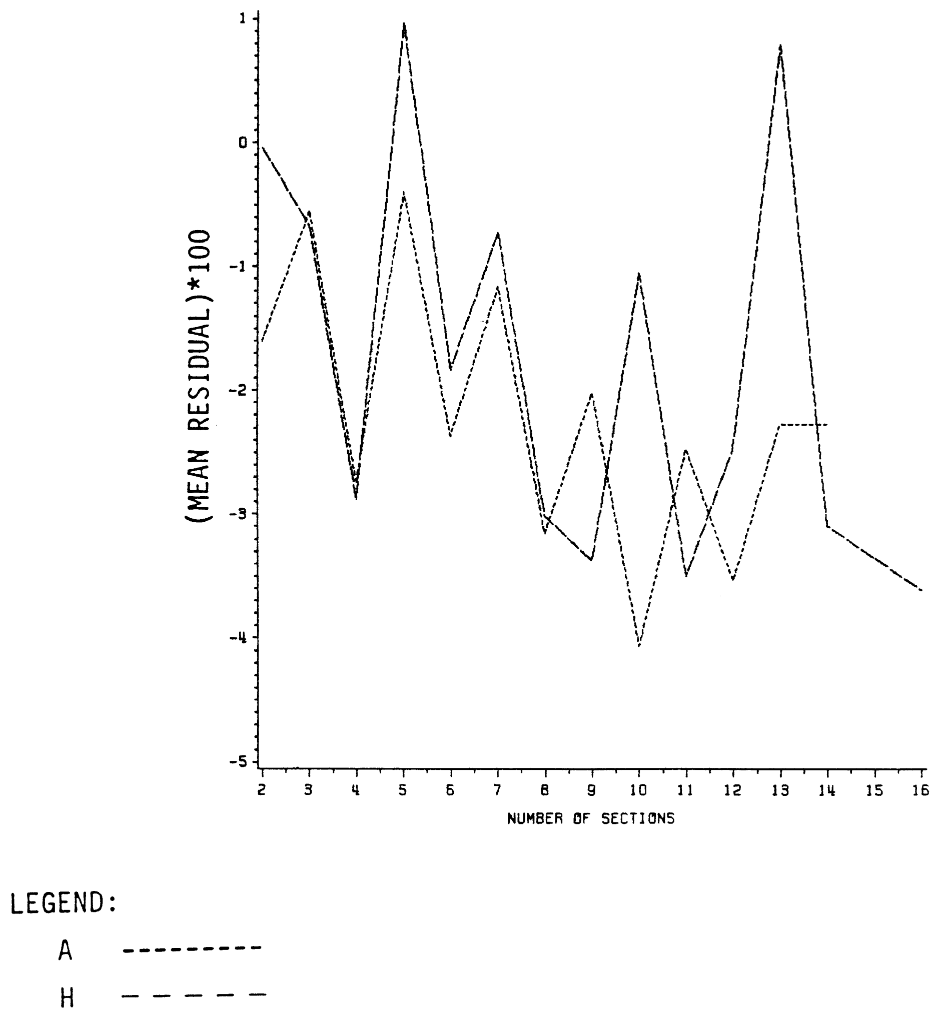


Figure 18. Mean percent residual versus number of sections used in updating for old-field loblolly pine plantation data set.

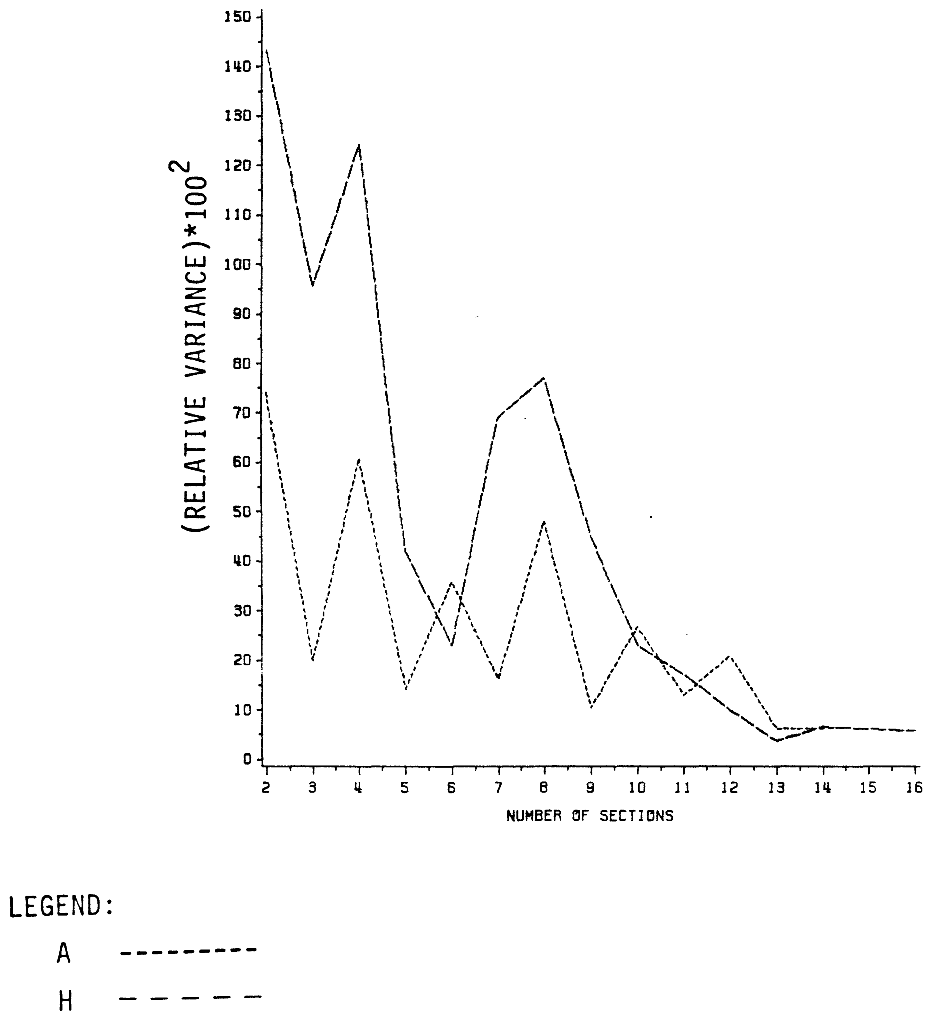
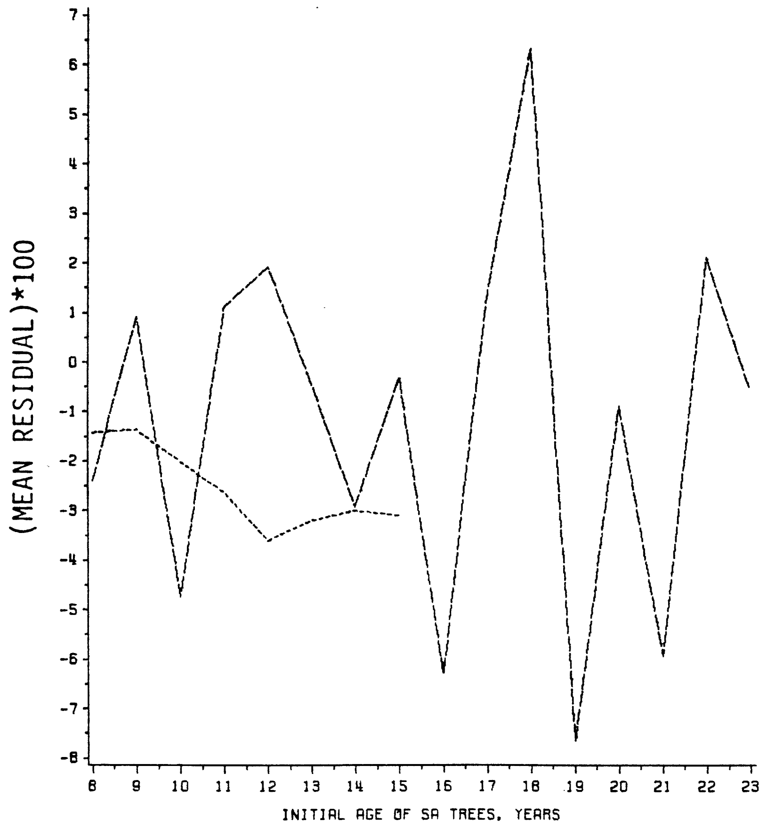


Figure 19. Variance of percent residual versus number of sections used in updating for old-field loblolly pine plantation data set.



LEGEND:

A ---------

H — — — — —

Figure 20. **Mean percent residual versus initial age of stem analysis sections used in updating for old-field loblolly pine data set.**

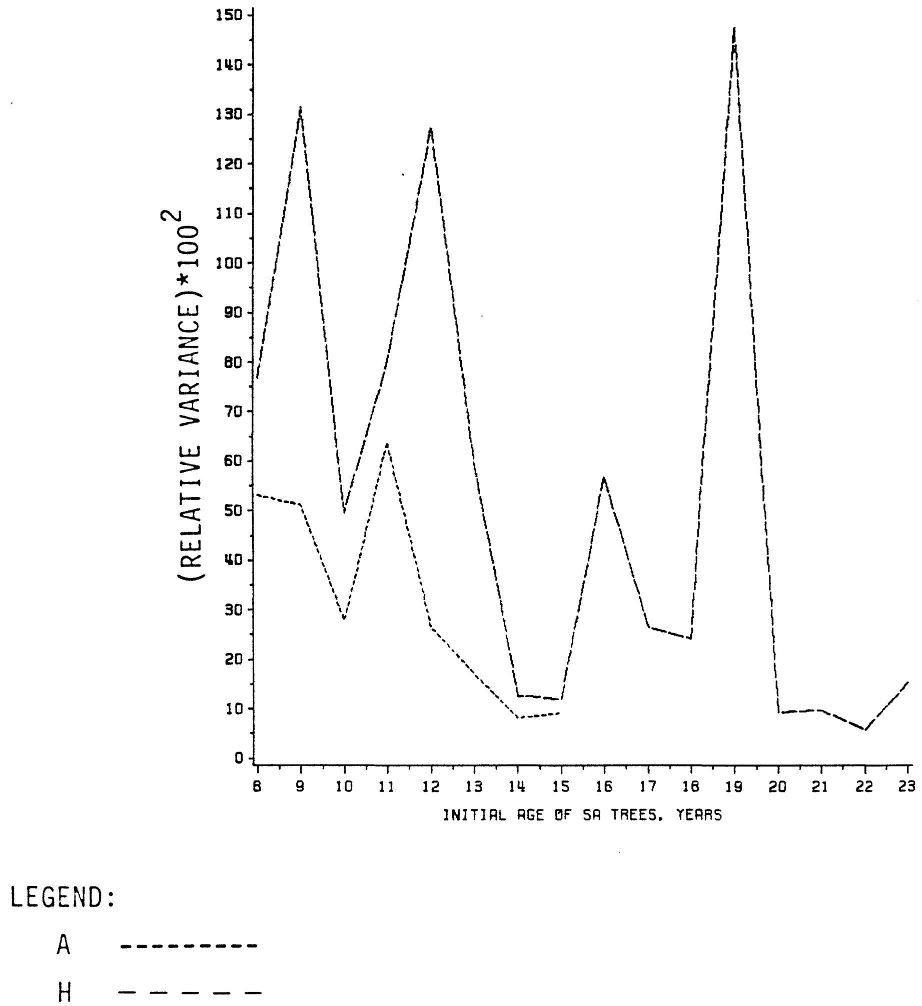


Figure 21. Variance of percent residual versus initial age of stem analysis sections used in updating for old-field loblolly pine data

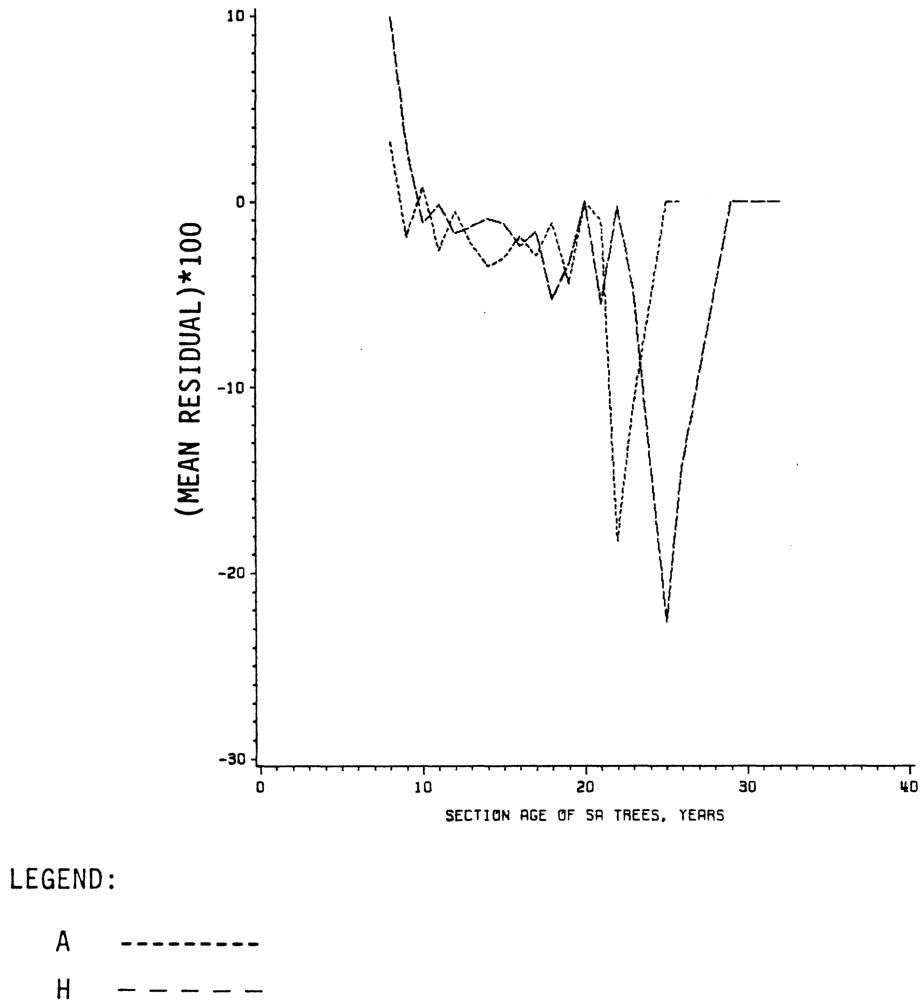


Figure 22. Mean percent residual versus age of sections used in updating for old-field loblolly pine plantation data.

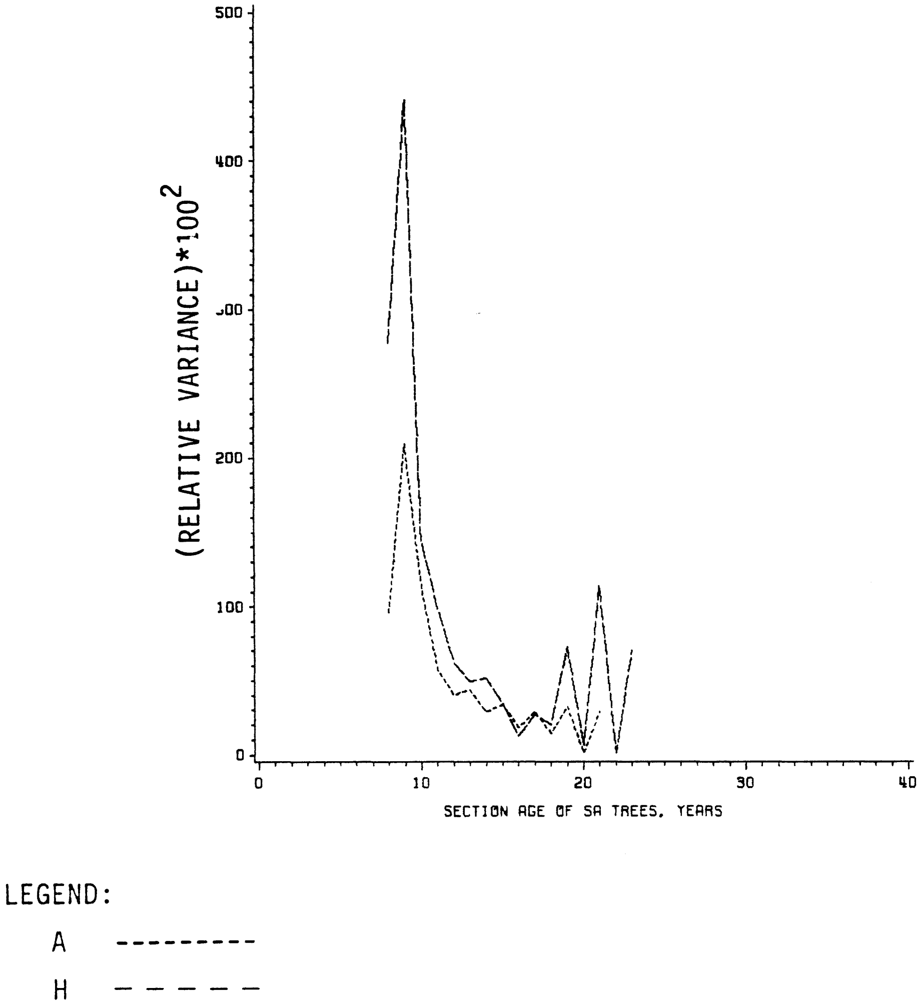


Figure 23. Variance of percent residual versus age of sections used in updating for old-field loblolly pine plantation data.

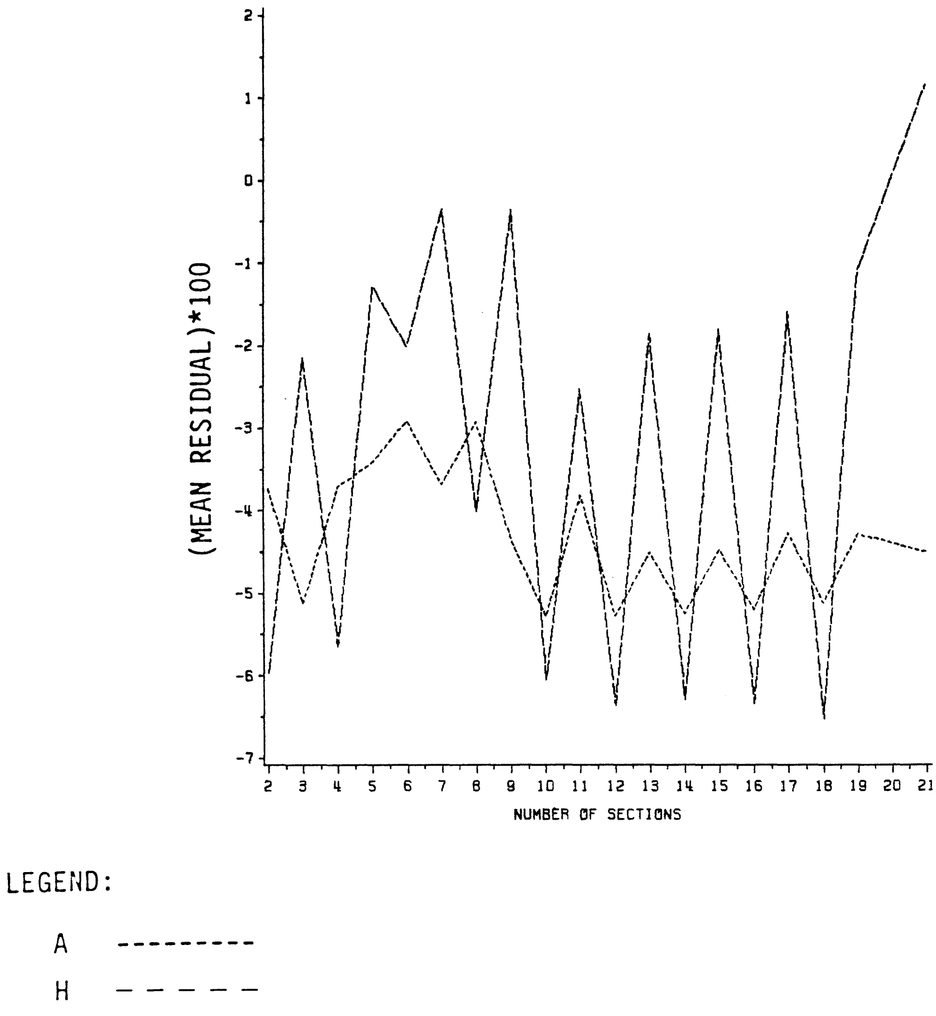


Figure 24. Mean percent residual versus number of sections used in updating for natural loblolly pine plantation data.

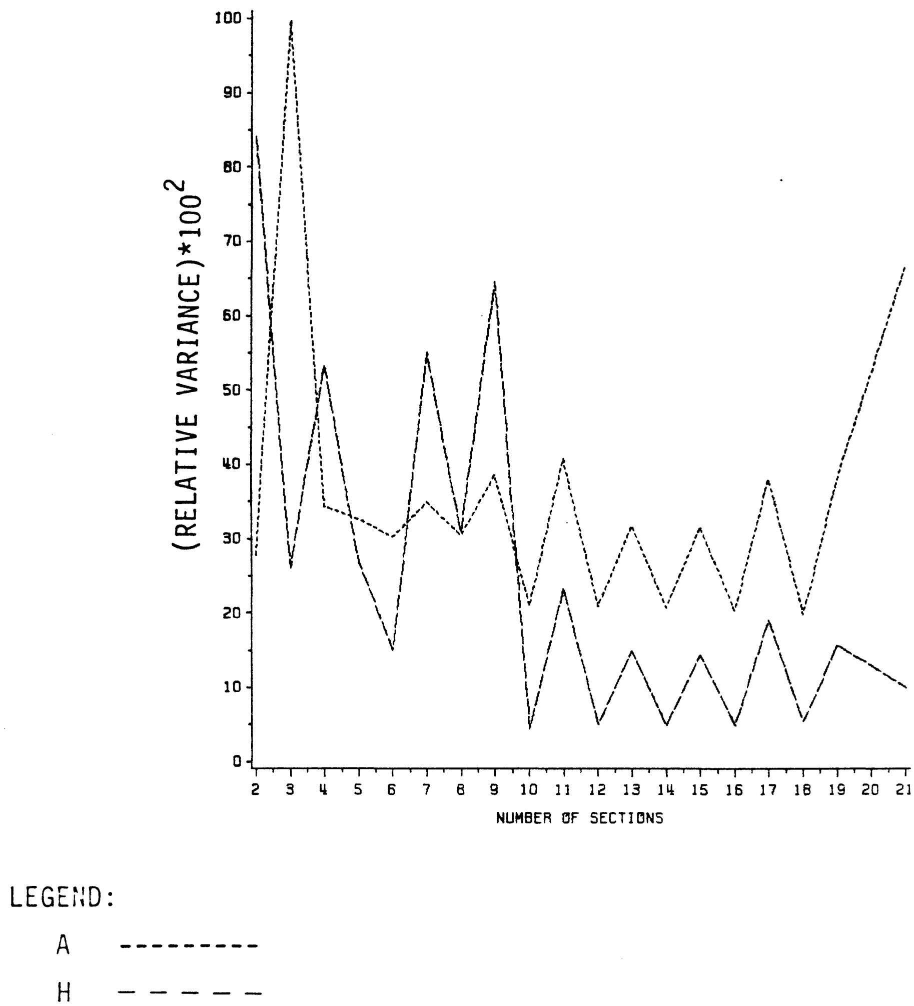


Figure 25. Variance of percent residual versus number of sections used in updating for natural loblolly pine stand data.

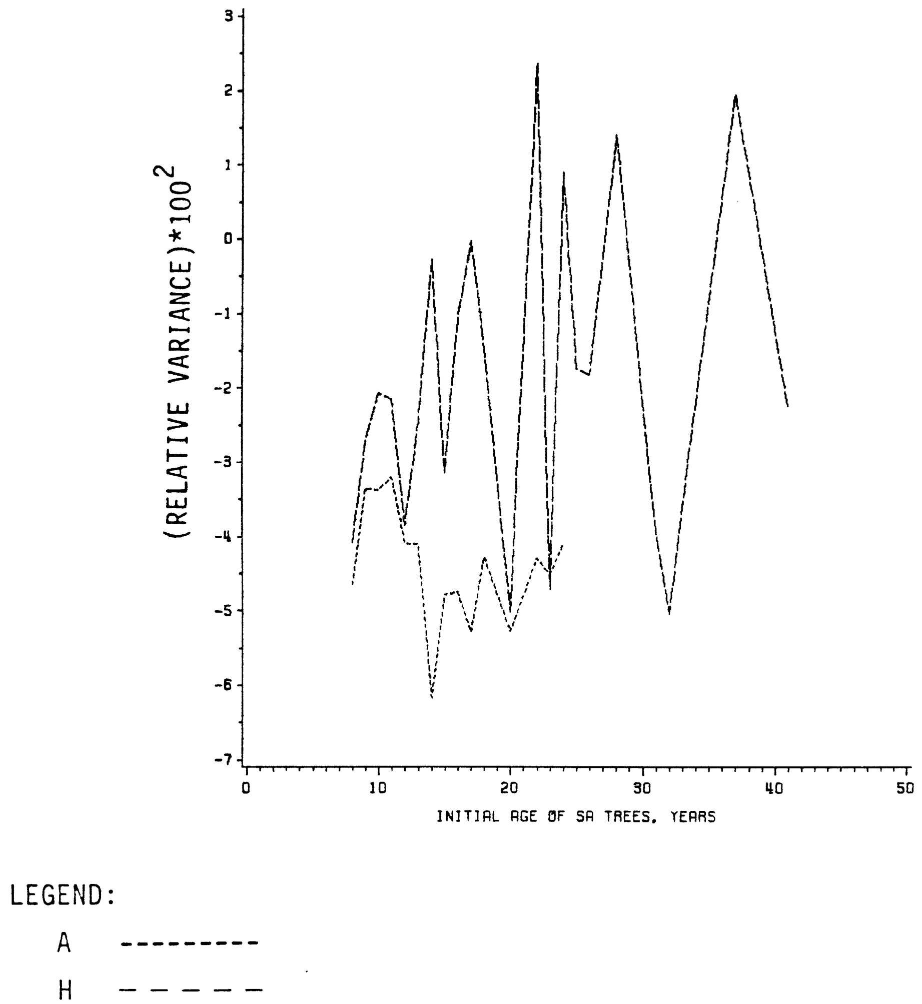


Figure 26. Mean percent residual versus initial age of stem analysis sections used in updating for natural loblolly pine stand data.

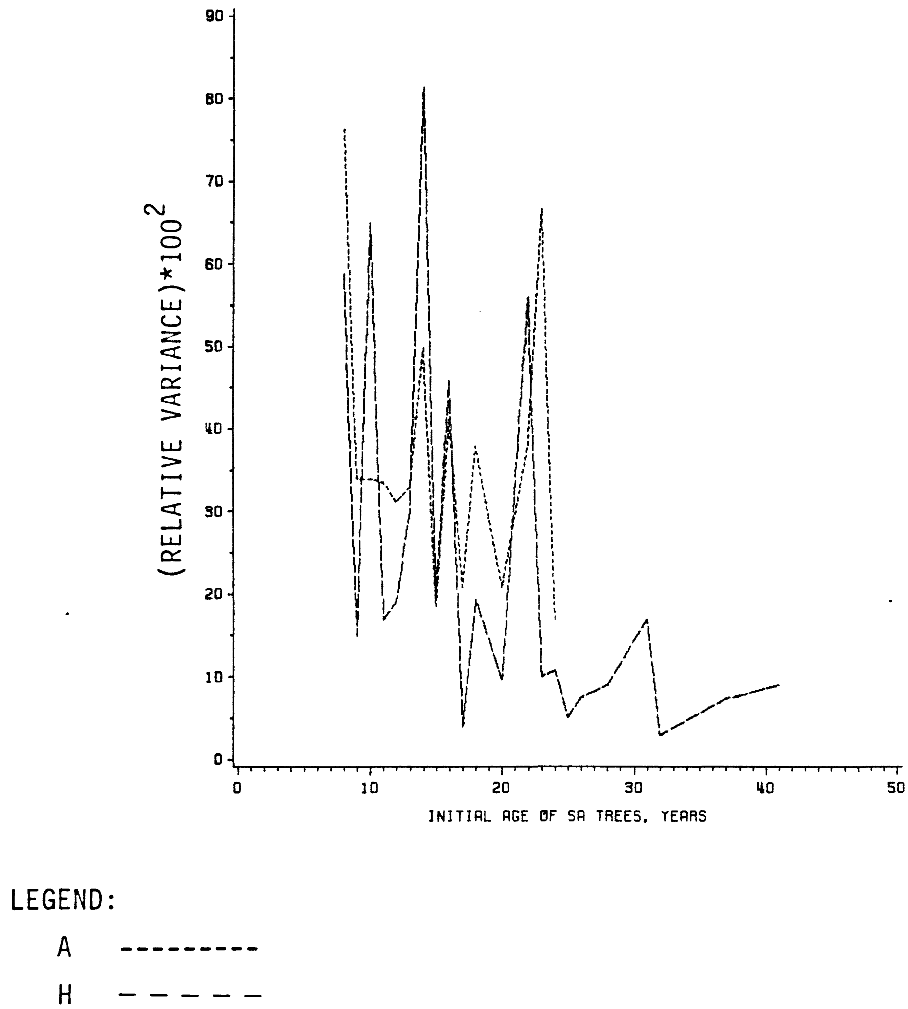


Figure 27. Variance of percent residual versus initial age of stem analysis sections used in updating for natural loblolly pine stand.

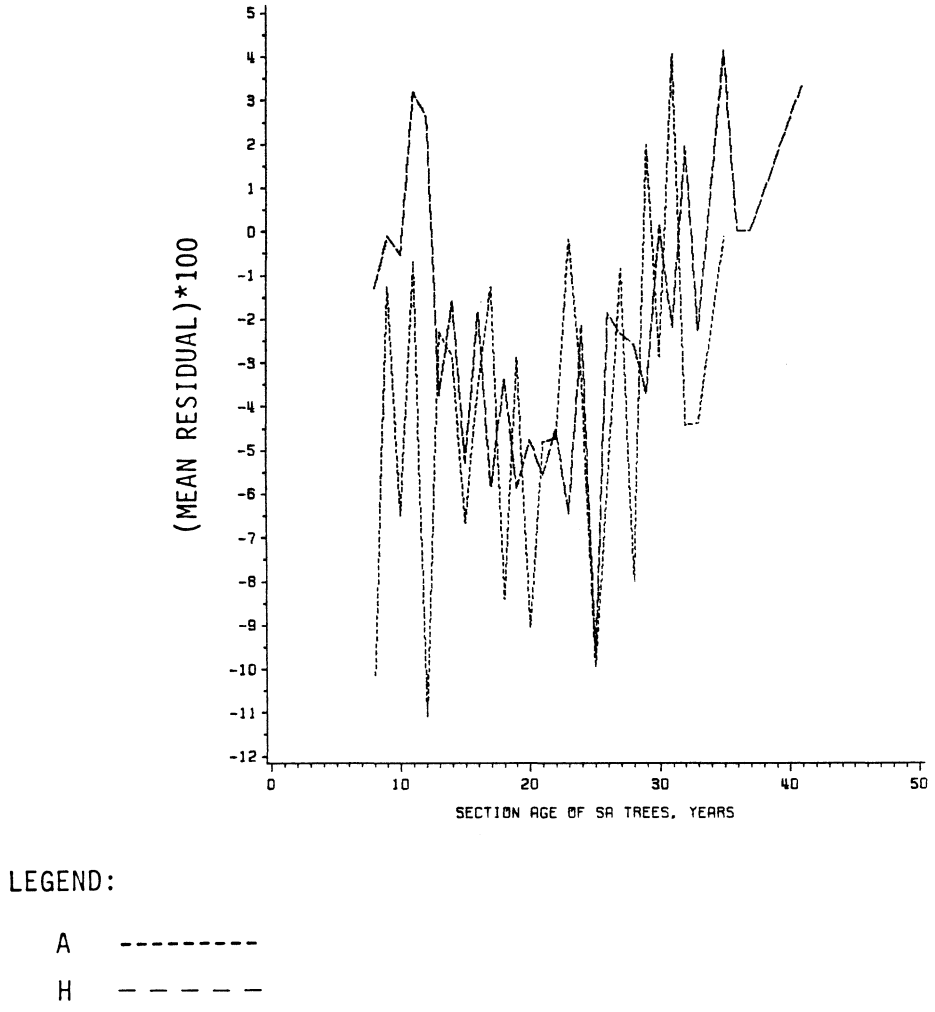


Figure 28. Mean percent residual versus age of sections used in updating for natural loblolly pine stand data.

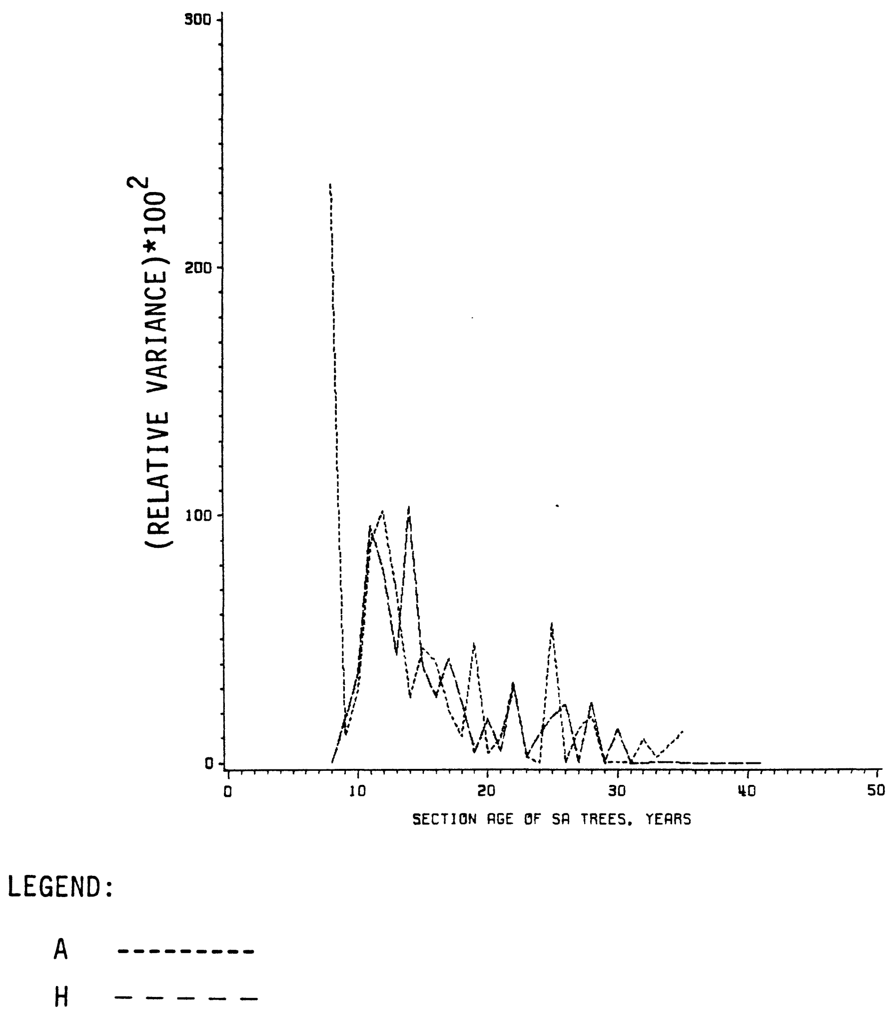


Figure 29. Variance of percent residual versus age of sections used in updating for natural loblolly pine stand data.

Table 13. **Summary statistics for updating models - Group 1, natural stand data**

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
K	- a/	- 11.2340 (331.26)	- 10.2470 (95.15)
L	-	- 11.4190 (345.31)	- 10.3990 (107.46)
M	-	- 11.4090 (345.43)	- 10.3920 (107.67)
N	-	8.8579 (198.09)	8.7254 (83.76)

a/ Values missing because there is no remeasurement data available from which this statistic can be calculated

Table 14. Summary statistics for updating models - Group 1, old-field data

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
K	- a/	11.1829 (640.57)	13.3557 (202.642)
L	-	11.1885 (640.80)	13.3865 (202.81)
M	-	11.1973 (641.11)	13.3950 (203.03)
N	-	6.5132 (290.70)	7.3972 (142.95)

Table 15. Summary statistics for updating models - Group 2, old-field data

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
A	- a/	- 2.013 (44.12)	- 1.5889 (23.36)
B	-	- 2.1137 (41.59)	- 1.7349 (21.53)
D	-	- 6.3213 (139.91)	- 6.6705 (66.53)
F	-	- .6447 (111.35)	- 1.3134 (71.54)
M	-	- 2.1282 (43.34)	- 1.6785 (23.16)
N	-	1.42629 (46.11)	2.5557 (25.58)

a/ Values missing because there is no remeasurement data available from which this statistic can be calculated

Table 16. **Summary statistics for updating models - Group 2, natural stand data**

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
A	- a/	- 4.0550 (41.98)	- 4.7396 (31.11)
B	-	- 3.0329 (38.13)	- 3.5309 (20.17)
D	-	- 3.4998 (97.02)	- 4.1875 (68.90)
F	-	- 1.6667 (83.38)	- 3.1469 (60.10)
M	-	- 4.2391 (49.42)	- 4.7551 (28.91)
N	-	3.8696 (50.45)	3.7328 (31.02)

a/ Values missing because there is no remeasurement data available from which this statistic can be calculated

Table 17. Summary statistics for updating models - Group 3, old-field data

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
H	- a/	- 1.1781 (82.11)	- .9746 (62.31)
I	-	- 2.8810 (56.46)	- 2.2752 (31.04)
J	-	.1311 (215.33)	- 1.1165 (140.47)
M	-	- 1.9707 (42.49)	- 1.3984 (22.23)
N	-	1.4263 (46.11)	2.5557 (25.58)

a/ Values missing because there is no remeasurement data available from which this statistic can be calculated

Table 18. Summary statistics for updating models - Group 3, natural stand data

Model	Overall Plot-Level (variance)	Overall Stem-Analysis (variance)	Mean Bias Across Plots (variance)
H	- a/	- 2.4255 (39.09)	- 3.3089 (33.65)
I	-	- 1.9560 (84.91)	- 2.4006 (69.46)
J	-	- 1.4238 (112.65)	- 3.4064 (84.43)
M	-	- 4.2391 (49.42)	- 4.7551 (28.91)
N	-	3.8696 (50.45)	3.7328 (31.02)

a/ Values missing because there is no remeasurement data available from which this statistic can be calculated

The vita has been removed from the scanned document