Development of a well-behaved site index equation: jack pine in north central Ontario

J.C.G. GOELZ

Southern Forest Experiment Station, Southern Hardwoods Laboratory, P. O. Box 227, Stoneville, MS 38776, U.S.A.

AND

T.E. BURK

Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, U.S.A.

Received July 19, 1991

Accepted January 13, 1992

GOELZ, J.C.G., and BURK, T.E. 1992. Development of a well-behaved site index equation: jack pine in north central Ontario. Can. J. For. Res. **22**: 776–784.

A base-age invariant site index equation for jack pine based on the Chapman–Richards function was produced that satisfied nine criteria of preferred behavior for site index equations. A difference form of the Chapman–Richards equation produced the best behavior; height equalled site index at base age, and the shape of the curves reflected the data. The data structure used to fit the difference equation was all possible differences rather than the conventional nonoverlapping sequential intervals because this improved the behavior of the model. Height-prediction equations typically use height at base age (site index) as a predictor variable. As site index is measured with error, the equation will be biased. This bias will be evident in the predicted height at base age and in the shape of the curves. Base-age invariant equations predict height and site index with the same equation and thus diminish the effect of stochastic predictor variables. The equation performed comparably to a previously published equation with a specific base age of 50 years.

GOELZ, J.C.G., et BURK, T.E. 1992. Development of a well-behaved site index equation: jack pine in north central Ontario. Can. J. For. Res. **22** : 776–784.

Une équation d'indice de fertilité indépendante de l'âge de référence a été établie pour le pin gris au moyen du modèle de Chapman–Richards. Cette équation satisfait neuf critères d'évaluation des équations d'indice de fertilité. Sous la forme de différences, l'équation de Chapman–Richards, démontre le meilleur comportement : la hauteur égale l'indice de fertilité à l'âge de référence et la forme des courbes correspond à celle des observations. Les données utilisées dans le calcul de l'équation des différences consistent en l'ensemble des différences possibles plutôt que l'approche usuelle des intervalles séquentiels; ce qui améliore le comportement du modèle. Les équations prédisant la hauteur utilisent habituellement la hauteur à l'âge de référence (indice de fertilité) comme variable explicative. Comme la mesure de l'indice de fertilité est entachée d'erreur, l'équation de la hauteur en est biaisée. Ce biais se manifeste dans les estimés de la hauteur à l'âge de référence et dans la forme des courbes. L'équation d'indice de fertilité indépendante de l'âge de référence prédit la hauteur et l'indice de fertilité dans la même équation, ce qui diminue l'effet des variables explicatives aléatoires. L'équation se compare favorablement avec une équation déjà publiée ayant 50 ans comme âge de référence.

[Traduit par la rédaction]

Introduction

Site index is the most used method of site quality evaluation for even-aged forest stands in North America (Carmean 1975). Site index may be defined as height of trees that have always been dominant or codominant and healthy at a predetermined age. Typically this age, referred to as base age, is set at 25, 50, or 100 years. The base age is specified so that it is somewhat less than rotation age. Often, breast-height age, rather than total age, is used because early height growth is erratic and often determined by factors other than site quality (Carmean 1975). Models involving site index include two processes: (*i*) estimating height at base age given height at some other age and (*ii*) estimating height at some desired age given height at base age. These two processes may be modeled by individual functions for each process, or by one equation that predicts height at any desired age, given a known height at any other age. In either case, a known height and age are used to predict height at some other age; thus site index equations are inherently of a difference form.

Desirable attributes of site index equations

Past site index equations have taken many forms. The various forms may have particular characteristics that are desirable and other characteristics that are undesirable. We present several characteristics that are desirable for site index equations:

Criterion 1: polymorphism

Early site index equations were typically anamorphic; the curve for a given site index was a constant proportion of the curve for another site index. Bull's (1931) early work shed doubt on the appropriateness of anamorphic curves, although it was largely ignored for 30 years. To allow polymorphism, Carmean (1972) broke his data into 10 ft (~3 m) wide classes of site index and fit individual equations for each class. However, for certain equation forms, polymorphism may be obtained by representing selected parameters as a function of site index. Clutter *et al.* (1983) presented two procedures that may provide polymorphism, the difference equation approach, and the parameter prediction approach. Almost all recently developed site index curves are polymorphic.

Criterion 2: inflection point

A sigmoid growth pattern is a paradigm of biology. Although some data sets may not show an inflection point, a function for general use should allow an inflection point.

Criterion 3: asymptote

Although the height of a tree increases annually until the leader is damaged or the tree is killed, an asymptote ensures that height is not unbounded at old ages. Ideally, the asymptote should be some function of site index such that the asymptote increases with increases of site index (Devan and

Burkhart 1982). However, a parameter representing an asymptote may be poorly estimated if heights at old ages are not included in the data set.

Criterion 4: logical behavior

In addition to the previous considerations, the equation should not allow unreasonable values for predicted height. Height should be zero at age zero (or breast height at breast-height age of zero), and height should be equal to site index at base age. If height decreases over any range of age, the equation form should be adjusted or discarded. Devan and Burkhart (1982) presented a site index curve that produced very large values of height at a young age; they discounted the importance, as the behavior was suitable for the range of ages that would be used in practice. Limits for usage must be appended to such curves; as the functions are continuous, any such limit is necessarily arbitrary.

Criterion 5: theoretical basis

By developing an equation from a theory, two benefits are obtained. The previous considerations may be explicitly incorporated into the equation form and the equation may perform better outside the data used to fit the equation (Pienaar and Turnbull 1973).

Criterion 6: base-age invariance

Base-age-specific equations predict height at some age using only site index and estimated parameters. If forest management for a given species was intensified and thus rotations were shorter than the specified base age, the utility of the curves would be diminished. Base-age invariant curves predict height at some age based on base age, height at base age, and the estimated parameters. Thus base-age invariant curves are more general, as they may predict height at any age given height at any other age. This attribute allows one function to predict site index, based on a measured height and age, as well as height, based on site index and age.

Some equations based on a specific base age may be solved for site index and others may not. The Chapman–Richards function (Chapman 1961; Richards 1959; Pienaar and Turnbull 1973; Ek 1971) is one of the most commonly used equations for site index curves; however, previous formulations of the Chapman–Richards model are impossible to solve for site index.

Separate equations for site index and height estimation will provide more efficient parameter estimates (Curtis *et al.* 1974). However, this is based on the assumption that it is appropriate to consider height or site index as measured without error when they occur on the right-hand side of an equation, but possess error when they are on the left-hand side of an equation. The consequences of that assumption have been largely unexplored in regard to site index methodology. Curtis *et al.* (1974) indicated that their findings were dependent on their concept of site index. We emphasize a system where the objective is to accurately predict height at any age, given height at any other age. This is somewhat different than the classical concept of site index.

The purpose of this paper is to present a base-age invariant form of the Chapman–Richards function that possesses the desirable attributes listed above. We have chosen the Chapman–Richards function as it is very flexible, is based on theory, and is almost the accepted paradigm for forest biometry. We compare our results to a previous model based on the same data set.

Development of a base-age invariant version of the Chapman–Richards function

In differential equation form, Chapman–Richards generalization of Bertalanffy's growth equation is

$$[1] \quad \frac{dH}{dA} = nH^m - kH$$

where H represents height, A represents age, and n, m, and k are parameters. The first term represents anabolic growth, and the second term represents catabolism; the equation is derived from a general theory of growth of organisms and well represents observed growth trends (Bertalanffy 1941; Richards 1959; Chapman 1961).

Equation 1 may be integrated to provide an equation for accumulated height rather than height growth. With the initial height and age specified as zero, the following integral form is obtained:

$$[2] \quad H = \left(\frac{n}{k}\right)^{\frac{1}{1-m}} \left[1 - \exp(-k(1-m)A)\right]^{\frac{1}{1-m}}$$

where variables are the same as in eq. 1.

Equation 2 may be simplified and written as

$$[3] \quad H = a[1 - \exp(-bA)]^c$$

where a, b, and c, are the respective functions of n, k, and m. Equation 3 was used by Lundren and Dolid (1970) to fit an equation to graphical site index curves. Carmean (1972) fit eq. 3 to individual site index classes. Brickell (1966) expanded eq. 3 by expressing each parameter as a function of site index. Several subsequent researchers have used other functions of site index to represent each parameter. Table 1 lists formulations used by various researchers. The resulting models expand eq. 3 to a function of 2 to 10 parameters. Burkhart and Tennent (1977) developed a model that requires only two parameters. They solved eq. 3 for a, then set height equal to site index and age equal to a base age of 20 years and inserted the equation for a back into eq. 3. Thus they explicitly rearranged eq. 3 to be a difference equation; this had the added benefit of insuring that height would equal site index at base age, a property that only the difference equation formulations possess. Newnham (1988) rearranged eq. 3 into a difference form by solving for b.

Beck (1971), Graney and Burkhart (1973), and Burkhart and Tennent (1977) found that the c parameter was not related to site index when eq. 3 was fit to individual trees and thus specified c to be common for all sites. Ek (1971) specified b to be common for all sites, as this reflected the graphical curves to which he wished to fit a function. Ek's (1971) model has been used by several other researchers (Hahn and Carmean 1982; Payandeh 1974*a*, 1974*b*; Carmean and Lenthall 1989; Carmean *et al.* 1989). Monserud (1984) found that Ek's model fit his data poorly. He modified the logistic equation to obtain a better fit. As the logistic equation is a special case of eq. 1, Monserud's findings indicate that a particular expansion of eq. 3 performed poorly; some unidentified expansion of eq. 3 exists that should fit as well as his logistic function.

The formulations in Table 1 expand the parameters of eq. 3 as functions of site index. Base-age invariant site index equations may be produced by predicting height at a given age

TABLE 1. Functions used to express the parameters of eq. 3 as functions of site index

Parameter	Function	Source
a	$b_1 S^{b_2}$	Ek 1971; Biging 1985; Newnham 1988
a	$b_1 + b_2 S$	Graney and Burkhart 1973; Beck 1971; Trousdell *et al.* 1974
a	$b_1 + b_2 S + b_3 S^2 + b_4 S^3$	Brickell 1966
a	$\dfrac{S}{1 - \exp(20 b_1 S)^{b_2}}$	Burkhart and Tennent 1977
b	b_3	Ek 1971; Biging 1985
b	$b_3 + b_4 S$	Graney and Burkhart 1973; Beck 1971
b	$b_1 S$	Burkhart and Tennent 1977
b	$\exp(b_5 + b_6 S + b_7 S^2)$	Brickell 1966
b	$b_3 + b_4 S + b_5 S^2$	Trousdell *et al.* 1974
b	$\ln\left(1 - \left(\dfrac{S}{b_1 S^{b_2}}\right)^{\frac{1}{b_3} S^{b_4}}\right)/50$	Newnham 1988
c	$b_4 S^{b_5}$	Ek 1971
c	$b_3 S^{b_4}$	Newnham 1988
c	b_5	Graney and Burkhart 1973; Beck 1971
c	b_2	Burkhart and Tennent 1977
c	b_4	Biging 1985
c	$\dfrac{1}{1 - (b_6 + b_7 S + b_8 S^{2)}}$	Trousdell *et al.* 1974
c	$\dfrac{1}{1 - b_8 S^{b_9} (\ln(S))^{b_{10}}}$	Brickell 1966

NOTE: Where site index (S) does not enter into the expression, the parameter was not expressed as a function of site index. Numbering of the parameters is sequential for a given source.

dependent on known height at some other age. Thus we sought to expand the parameters of either eq. 2 or 3, or some intermediate form, as functions of height and age.

We considered two very general functions to expand the parameters of eq. 2 or 3 as functions of height and age. Hoerl's special functions (Daniel and Wood 1980) produces a family of curves that is very flexible and is well suited to simplification if one of the parameters is not significant; we expanded Hoerl's special functions to yield

$$[4] \quad Y = b_1 X_1^{b_2} \exp(b_3 X_1) X_2^{b_4}$$

where Y represents any parameter of eq. 2 or 3, X_i represents height or age, in either order, and the b_i are parameters. As the parameters of eq. 2 should be positive values less than one, we proposed using the logistic function to describe them:

$$[5] \quad Y = \frac{1}{1 + \exp[f(X_1, X_2)]}$$

where Y represents any parameter of eq. 2 and f represents some function of age and height.

For a given tree, eq. 3 would thus be expanded as follows:

$$[6] \quad H_i = f_1(H_{(i)} A_{(i)}) \ [1 - \exp(-f_2(H_{(i)} A_{(i)}) A_i)]^{f_3(H_{(i)} A_{(i)})}$$

where f_1 is a function representing parameter a, f_2 is a function representing parameter b, f_3 is a function representing c, $H_{(i)}$ represents any height other than H_i, and $A_{(i)}$ represents any age other then A_i. Equation 2 was expanded similarly.

In addition to the six criteria we listed before, we considered three additional desirable attributes for the behavior of the functions describing the parameters:

Criterion 7

Each height–age pair produces a curve. Any given point along that curve could be used to generate another curve. To obtain base-age invariance, all height–age pairs should generate identical or similar curves. For a given series of height and age observations that follow the function without error, the estimated parameters should be relatively constant for all ages. Equations 2 and 3 may be solved for a given parameter, thus ensuring this behavior for that specific parameter; however, difference equations do not obtain such behavior for all parameters.

Criterion 8

If 50 years is substituted into $A_{(i)}$ and site index is substituted into $H_{(i)}$, the parameters should approximate those obtained from a fit of a base-age-specific version of a function fit to the same data, or the parameters should reflect the estimates from fits to individual plot data.

Criterion 9: parsimony

The final form of the functions describing the parameters should possess few terms.

Our emphasis throughout this work was the development of a site index system that possessed the behavior described by the nine criteria we listed. Initial screenings of models were based on the behavior of the model. Fit statistics were used for comparison only after appropriate behavior was ensured. The adequacy of our final model was based on comparisons to a previously published base-age-specific model with regard to the behavior of both the height and site index prediction equations.

The data

Data collection is discussed fully in Carmean and Lenthall (1989). There were 141 plots located in north central Ontario, which were subjectively selected to represent the range of site quality and soils in the area. Three to five trees were destructively sampled at each location. The trees were sectioned at the stump, at 0.75, 1.3, and 2.0 m, and 1-m intervals to 13 m, and at 0.5-m intervals thereafter. Age at each section height was determined in the laboratory. Carmean's (1972) correction was used to estimate actual height for each age. Dyer and Bailey (1987) found that Carmean's method worked best among several other methods. Newberry (1991) noted that Carmean's correction is not appropriate for the topmost section of the tree. However, whorls or annual bud scale scars provide exact heights for recent years. The individual trees in a plot were averaged into a single series of height–age pairs for a given plot. The plot average was used to provide height measurements at 5-year increments above breast height. There were 109 plots used for estimation, leaving 32 plots for verification of any model fitted. Estimation plots were randomly chosen within classes of site index.

Analysis

Initially, data for individual plots were fitted to eqs. 2 and 3. This was carried out to identify relationships between the parameters and height at a given age. The n parameter exhibited a strong linear relationship with height at any given age; the relationship was so strong that n could be considered as an index of productivity (cf. Hamlin and Leary 1987). This strong relationship led us to initially consider eq. 2 with parameters modeled by eq. 4. Starting values for nonlinear regression were determined from linear estimation of eq. 4 using the parameters estimated from fitting individual plots as the dependent data.

Although numerous modifications were attempted, we could not fit a height-growth equation that provided parameter estimates that were close to the estimates from fitting individual plots. The asymptote was much larger than expected. The equations were invariably ill-behaved with regard to the criteria we had set forth.

Initial attempts at fitting an expanded version of eq. 2 used the previous measurement of height and age as a predictor of height at a given age. This is the typical procedure for fitting equations that are inherently of a difference form (Clutter *et al.* 1983; Furnival *et al.* 1990). We hypothesized that the inflated asymptotes were a consequence of using only heights immediately preceding the predicted heights. Thus we next considered using all height measurements of a given plot, other than the dependent variable, as predictor variables. Rather than expanding the equation to allow nine or more other height and age measurements as predictor variables, we expanded the number of observations to allow prediction by all other measurements for a plot. A tree with 10 measurements would thus provide 90 observations, each height predicted by nine heights at other ages. When we used the data organized in this way, the parameter estimates were relatively consistent with estimates for individual plots. We recognized that this would introduce a lack of independence among observations; we modeled the error structure as detailed later.

Although parameter estimates were close to the fits to individual plots, the resulting curves did not fulfill the desired criteria. Specifically, for a sequence of height–age pairs that followed the Chapman–Richards function without error, the parameters, when solved to the form of eq. 2 or 3, varied considerably with age. This observation was in violation of criterion 7. By solving for one of the parameters in eq. 2 or 3 and replacing the solution into the equation, we could eliminate the problem for one parameter and explicitly produce a difference equation. Burkhart and Tennent (1977) and Clutter *et al.* (1983) suggested solving for a, while Newnham (1988) suggested solving for b in eq. 3. We considered solving for a, b, and c in eq. 3, solving for n in eq. 2, and solving for n, b, and c in forms intermediate between eqs. 2 and 3. Except when n or a were considered, the equations (*i*) converged to poor estimates, (*ii*) did not converge within 100 iterations, or (*iii*) did not solve the problem. Thus our subsequent efforts were based on a difference form after solving for n or a.

To simplify the functions describing the parameters, we needed to make hypothesis tests regarding the parameters of eq. 4. To obtain more efficient estimates of errors, autocorrelation was considered. The ordinary nonlinear least squares model is written

$$[7] \quad Y_i = f(X_i, \beta) + e_i$$

where the error terms are independent and identically distributed. The error term may be expanded to allow first-order autocorrelation:

$$[8] \quad e_i = \rho\varepsilon_{i-1} + \varepsilon_i$$

where the ε_i are now independent and identically distributed. As we used all possible growth intervals, our model is slightly more complex:

$$[9] \quad Y_{ij} = f(X_i, Y_j, X_j, \beta) + e_{ij}$$

where Y_{ij} represents prediction of height i by using Y_j (height j), X_i (age i), and X_j (age $j \neq i$) as a predictor variables. Similarly, the error term must be expanded:

$$[10] \quad e_{ij} = \rho\varepsilon_{i-1,\,j} + \gamma\varepsilon_{i,\,j-1} + \varepsilon_{ij}$$

The parameter ρ represents the autocorrelation between the current residual and the residual from estimating Y_{i-1} using Y_j as a predictor variable. The parameter γ represents the relationship between the current residual and the residual from estimating Y_i using Y_{j-1} as a predictor variable.

Generalized least squared weights observations according to their variance. The autocorrelation parameters vary the weight of each observation by reducing the residual proportional to a previous residual. However, when either the first observation for a tree is used as a predicted or predictor variable, there is no appropriate previous observation. However, the following weighting, derived simply from the model of first-order autocorrelation (Wonnacott and Wonnacott 1976), is appropriate when i and j are the first observation for a tree

$$[11] \quad \omega_{ij} = \sqrt{(1 - \rho^2)\ (1 - \gamma^2)}$$

when only i is the first observation for a tree

$$[12] \quad \omega_{ij} = \sqrt{1 - \rho^2}$$

and when only j is the first observation for a tree

$$[13] \quad \omega_{ij} = \sqrt{1 - \gamma^2}$$

TABLE 2. Parameter estimates and standard errors for eqs. 10 and 16

Parameter	Estimate	Standard error
b_1	0.0185	0.0005
b_2	1.3382	0.0218
b_3	0.4257	0.0111
b_4	1.0464	0.0036
ρ	0.3975	0.0227
γ	0.5721	0.0228

We fit several models specifying ρ, γ, or both ρ and γ as zero. The parameter estimates, with the exception of the autocorrelation parameters, were insensitive to the specified error structure. However, the correct error structure allows correct hypothesis tests regarding significance of the parameters.

Heteroscedasticity was also apparent during fitting attempts. Rather than use empirical weighting, we observed a pattern that was described well by the following weighting function:

$$[14] \quad \bar{\omega}_{ij} = \min\left(w_{ij}, \frac{1}{w_{ij}}\right)$$

where

$$[15] \quad w_{ij} = \left(\frac{1 - e^{-bX_1}}{1 - e^{-bX_2}}\right)^c$$

The above formula assumes use of eq. 3; a similar formula could be given for eq. 2. In eq. 15, b and c could be expanded as functions of predictor height and age; X_1 represents predictor age and X_2 represents age at predicted height. The function weights proportionally to the ratio of predicted height at X_1 to predicted height at X_2.

Results and discussion

Description of the final model

Although approximately 100 models were fit, almost all lacked the appropriate behavior. One model was selected based on its superiority in regard to bias and root mean squared error for both calibration and validation data and in regard to bias for subsets of age and height. The form of the selected model is

$$[16] \quad \hat{H}_2 = 1.3 + (H_1 - 1.3)\left(\frac{1 - \exp\left(-b_1\left(\frac{H_1}{A_1}\right)^{b_2} A_1^{b_3} A_2\right)}{1 - \exp\left(-b_1\left(\frac{H_1}{A_1}\right)^{b_2} A_1^{b_3} A_1\right)}\right)^{b_4}$$

Equation 16 is the difference form of eq. 3 based on solving for parameter a. H_1 and A_1 represent the predictor height and age, respectively; $\hat{H}_2$ represents the predicted height at age A_2, which may be greater or less than A_1. The b parameter of eq. 3 was expanded as a function of the predictor height and age. A_1 appears twice in the function representing b. Although this is largely an artifact of the fitting process (a preliminary form included the H_1/A_1 term) we did not simplify because eq. 16 converged faster. The parameter estimates and standard errors of the estimates are presented in Table 2.

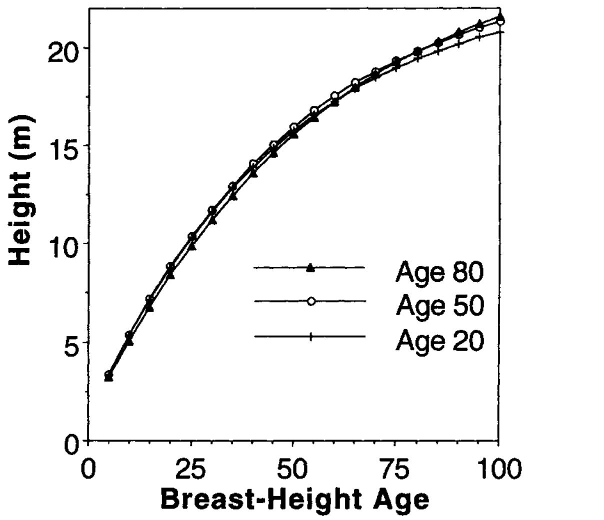

FIG. 1. Height-growth curves using three different base ages. The base-age-50 line is for a height of 16 m at base age. The base-age-20 and the base-age-80 lines intersect the base-age-50 line at their respective base ages.

Equation 16 was not constrained to produce the behavior described by criterion 7. The degree of similarity is expressed in Fig. 1. The height–age pair 16 m – 50 years breast-height age was used to generate a curve. From this curve, heights at 20 years breast-height age (8.83 m) and at 80 years breast-height age (19.82 m) were also used to generate curves; the range of age from 20 to 80 represents the range of stand breast-height age that would likely be applied to site index equations for jack pine in north central Ontario. In essence, we are using three different base ages to generate curves. Although there is a systematic pattern (the base-age-20 curve is slightly less than the base-age-50 curve for older ages and the base-age-80 curve is less than the base-age-50 curve for young ages) the discrepancy is never greater than 0.5 m.

Our difficulties in fitting a base-age invariant form of the Chapman–Richards function reflected those found by Monserud (1984) in fitting a base-age-specific form. Similarly, Cieszewski and Bella (1989) report an inability to produce consistent height-growth and site index curves from the Chapman–Richards equation and several related functions. However, we were eventually able to produce a model that fulfilled our criteria. Equation 16 indicates that the c parameter of eq. 3 does not depend on site quality. This agrees with Graney and Burkhart (1973), Beck (1971), and Burkhart and Tennent (1977). The results of fitting eq. 2 to data for individual plots indicated that n was very strongly related to height at a given age and thus possibly was indicative of site quality. Unfortunately, subsequent analysis indicated that a difference equation form was required that eliminated the asymptote parameter.

The data structure used to fit a difference equation may greatly affect parameter values. We found that the best data structure for fitting this difference equation was all possible growth intervals. Borders *et al.* (1988) considered three possible data structures for fitting a basal area growth model: (*i*) nonoverlapping sequential intervals, (*ii*) all possible

TABLE 3. Error criteria for eq. 16 and for Carmean and Lenthall's (1989) model

	Eq. 16	Carmean and Lenthall
Calibration data		
All data		
Bias	0.132	
MAD	0.764	
RMSE	1.149	
$A_1 = 50$		
BiasH	−0.179	0.007
MADH	0.551	0.544
RMSEH	0.777	0.722
$A_2 = 50$		
BiasS	0.413	0.004
MADS	0.856	0.869
RMSES	1.295	1.218
Validation data		
All data		
Bias	0.186	
MAD	0.647	
RMSE	0.933	
$A_1 = 50$		
BiasH	−0.110	0.129
MADH	0.443	0.478
RMSEH	0.606	0.621
$A_2 = 50$		
BiasS	0.410	0.033
MADS	0.754	0.751
RMSES	1.106	1.076

NOTE: Bias, mean absolute deviation (MAD), and root mean squared error (RMSE) are presented for the calibration data (109 plots) and the validation data (32 plots). The values are given for three different situations: (*i*) all data, which represents all possible combinations of using heights as predictor and predicted variables; (*ii*) $A_1 = 50$, which represents using height at age 50 as a predictor variable; (*iii*) $A_2 = 50$, which represents using height at age 50 as the predicted variable. The error criteria could not be calculated for all data with Carmean and Lenthall's model, as it uses a base age of 50.

growth intervals, and (*iii*) longest growth intervals. They found modest differences in their parameters depending on the data organization. Although the data of all possible growth intervals were not superior, with regard to prediction of a validation data set, the result is specific to their data and models. We intend to conduct a simulation study to investigate bias and efficiency of parameter estimation under several alternative data structures. Furnival *et al.* (1990) demonstrated that using all possible differences is identical with an analysis of covariance method and a method of weighted parameter prediction for linear models.

Comparison to a published base-age-50 system for site index

Error criteria

Carmean and Lenthall (1989) used the same data to develop a base-age-50 site index system. The curves generated by eq. 16 are presented in Fig. 2, along with the corresponding height-growth curves of Carmean and Lenthall (1989). In Table 3 we list error criteria for our model and for Carmean and Lenthall's (1989) model. For the data used in fitting, Carmean and Lenthall's (1989) model is slightly superior for mean absolute deviation and root mean squared error. Carmean and Lenthall's model is considerably superior with regard to bias.

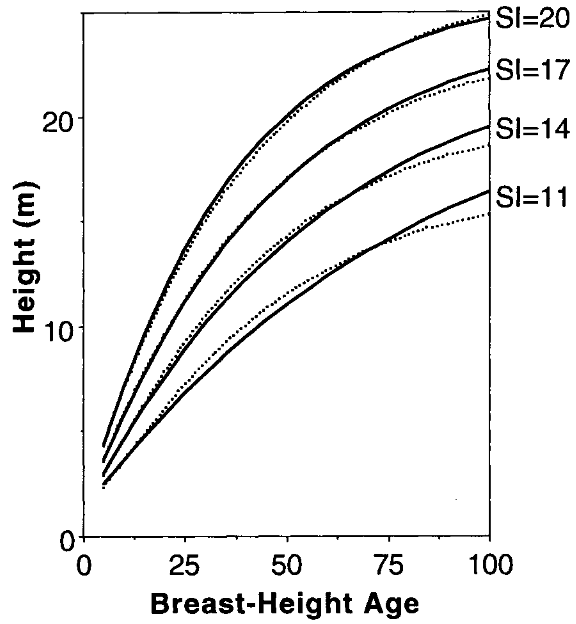

FIG. 2. Height-growth curves for jack pine at four values of site index (SI). The solid lines are generated with the base-age invariant curves from this study, with base age set at 50 years. The dotted lines are from the base-age-specific curves of Carmean and Lenthall (1989).

The same error criteria were calculated for the validation data. Equation 16 was better than Carmean and Lenthall's (1989) height-growth model ($A_1 = 50$) with regard to all three error criteria. Carmean and Lenthall's (1989) site index prediction model ($A_2 = 50$) has considerably lower bias and was slightly better for the other error criteria.

These error criteria indicate that eq. 16 produces a mean absolute deviation and root mean squared error that is about the same as, or slightly larger than, the corresponding equation presented in Carmean and Lenthall (1989). The equations of Carmean and Lenthall have much lower bias. Two factors provide an advantage to base-age-specific equations over a base-age invariant equation fit by weighted least squares in these comparisons based on unweighted residuals: (*i*) weighted least squares will consistently produce unweighted residuals that are biased and with a greater mean squared error than unweighted least squares (we compared residuals derived from an inappropriate error assumption with those from a proper assumption, but compared them on the basis of the inappropriate structure) and (*ii*) comparisons were necessarily made using a base age of 50; the base-age-specific equations minimized residual squared error for this age, and the base-age invariant equations minimized residual squared error for all possible base ages. These factors minimize the observation of greater bias although it may be important in some instance. Thus, neither system for site index is conclusively superior for a base age of 50. As eq. 16 may predict height at any age, given height at any age, it has greater capabilities.

Curve shape

The shape of the height over age curves vary according to some pattern. Individual curves for each unique site index could be fit; this would provide the best fit for each site index; it would recover all of the variation of shape across the range

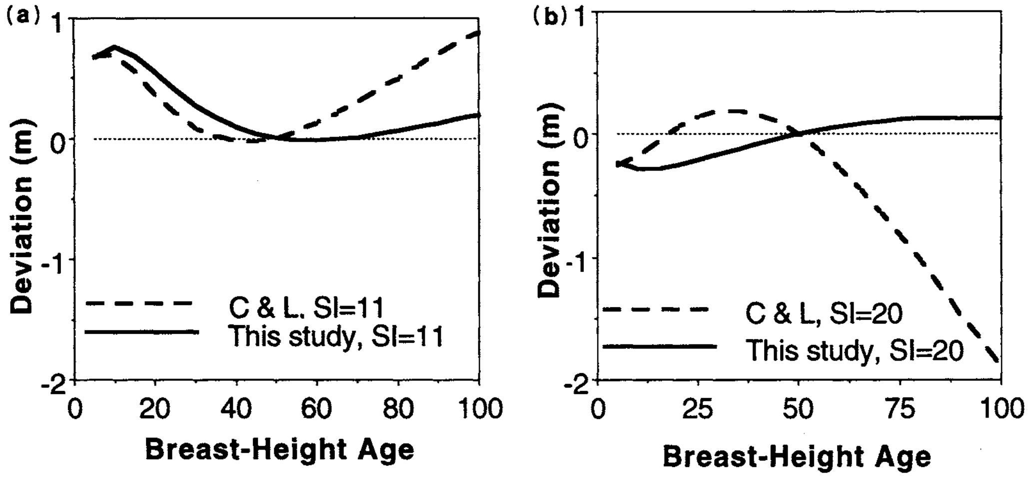

FIG. 3. Deviations between height-growth equations fit to individual classes of site index (SI) and polymorphic curves fit to all data simultaneously. The solid line is for the function fit in this study, and the broken line is for the function of Carmean and Lenthall (C & L) (1989).

of site indices. Alternatively, individuals could be classified into several site index classes and functions fit for each class (Carmean 1972); this would recover most, but not all, of the variation across site index. The narrower the classes, the greater would be the degree of resolution. At the extreme of pooling all data, the curve shape would be identical for all site indices; the curves would be anamorphic. By expressing one or more parameters of a height-growth equation as a function of site index, the curve shape will vary across the range of site indices. The changing pattern across site index will necessarily be continuous, but ideally most of the variation in curve shape will be recovered.

We sought to compare eq. 16 with Carmean and Lenthall's (1989) height-growth equation with regard to how well they recover the change of shape across site index. Lenthall (1986) fit the Chapman–Richards function (eq. 3) to the same data for individual site index classes (9-, 11-, 14-, 16-, 18-, and 20-m classes). Presumably, the equations for these individual classes will represent the true curve shape better than a single equation fit for all values of site index. As the 9-m class was based on only four plots, we chose to compare the 11- and 20-m site index class equations to the base-age-specific curve of Carmean and Lenthall (1989) and our results for eq. 16. In Fig. 3, the deviations from the equations for the individual classes are presented for Carmean and Lenthall (1989) and eq. 16. For site index 11, Carmean and Lenthall's curve (1989) performs slightly better before age 50, but diverges afterward. For site index 20, eq. 16 recovers the true shape of the curves very well, while Carmean and Lenthall's (1989) equation underestimates height at older ages. Equation 16 is better at recovering the true change of shape in response to site quality.

"Regression towards the mean" (Wittink 1988) implies that the slope of the regression line is less than the slope of the major axis of the data. This property follows directly from the assumption of fixed predictors. In nonlinear regression the predicted values also exhibit regression towards the mean; however, the effect is not simply a function of a linear slope parameter, but is additionally expressed as a change in the shape of the curve. As site index increases, the shape of the curve changes from a relatively linear, gradual approach to the asymptote to a more strongly curvilinear, more rapid approach to the asymptote. As shown in Figs. 2 and 3, the base-age-50 curves underestimate this trend. The site index 20 curves for Carmean and Lenthall (1989) do not flatten out as abruptly as they should, and the site index 11 curves flatten out too much. In Fig. 3, the curves for Carmean and Lenthall (1989) are corrected to pass through site index at base age, although in Fig. 2 they are not. Regression towards the mean was also evident in the more typical sense; the equations of Carmean and Lenthall (1989) slightly underestimated height at base age for higher values of site index and slightly overestimated height at base age for lower values of site index (Fig. 2).

Regression towards the mean was also evident in the results of the studies included in Table 1. Insufficient information was available to assess whether the studies produced equations whose shape was incorrect. However, with the exception of the difference equations and the results of Biging (1985), the equation fit in these studies consistently underestimated height at base age for higher site index and overestimated height at base age for lower site index.

As a substitute for linear regression, geometric mean regression (Ricker 1973) may be used when the predictor variables are stochastic. As an alternative to minimizing least squares, we considered minimizing:

$$[17] \quad \text{loss} = \text{abs}(H_1 - \hat{H}_1)\,\text{abs}(H_2 - \hat{H}_2)$$

where the predicted heights are determined by current parameter values. This is analogous to the geometric mean regression for simple linear regression. This procedure did not provide the appropriate behavior for the models we tested. We considered several variations of eq. 17, but found similar results.

Northway (1985) considered the problem of fitting base-age specific site index equations while recognizing that site index is a stochastic predictor variable. Given current parameter values and a number of height–age pairs for a tree, Northway suggested using the average of the predicted site index for that tree as the independent variable and then

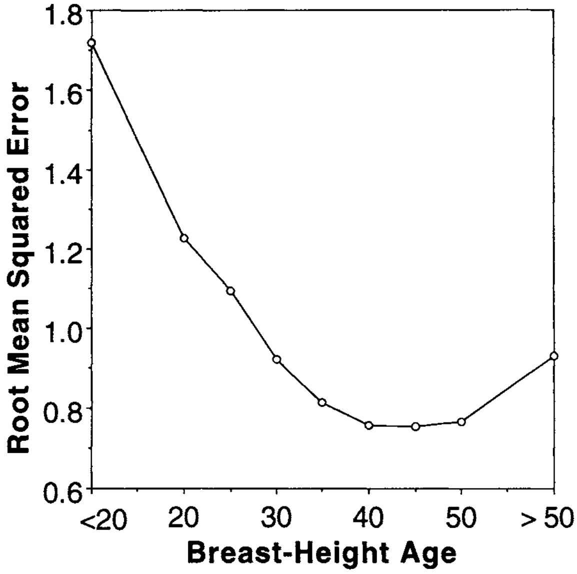

FIG. 4. Root mean squared error of height estimation related to choice of base age. The minimum occurs at age 45.

updating after each interation. Northway's (1985) suggestions can be considered a type of nonoptimal instrumental variable regresson (Walters *et al.* 1989), where the instrument, predicted site index, is not independent of the errors. From another prespective, Northway's method is also similar to two-stage least squares, where the functions may be nonlinear and the endogenous variables are used to estimate the instruments; the two stages are repeated at each iteration of fitting. Thus, it could be shown that Northway's method is not optimal from this perspective as well. We attempted a procedure similar to Northway's method; however, it was slow to converge and did not predict as well as alternative models, particularly for the extremes of site index.

While we do not provide a method that explicitly considers stochastic predictor variables, our observations from fitting numerous models indicates that base-age invariant difference equation models minimize the effect of stochastic predictor variables. Specifically, difference equations assure that predicted height equals site index at base age and, for our models, follow observed paterns of height growth better than other forms. Fitting base-age invariant models is analogous to fitting Y on X and X on Y equations simultaneously and thus can minimize the effect of stochastic predictor variables. We believe that using all possible growth intervals, rather than using one height as a predictor variable, may also improve recovery of the observed patterns of height growth.

Choice of a base age

The base age for site index equations should be selected according to the following considerations: (*i*) The base age should be less than or equal to the youngest rotation age under typical management. Presumably, stands older than the rotation age will be uncommon after a forest has been regulated. (*ii*) The base age should be close to the rotation age, as site index is mainly required for estimation of yield, and yield at final harvest is most important and should be predicted most precisely. (*iii*) The base age should be chosen such that it is a reliable predictor of height at other ages. Height at young ages may not reliably predict height at older ages; conversely, if the height-growth curves for different site indices converge at older ages, a somewhat younger age may best differentiate site quality. In northern Ontario, most jack pine stands are harvested between 50 and 80 years; more intensive management (harvesting smaller products, thinning, initial spacing) could shorten the rotation.

To address the third consideration, we calculated root mean squared error from the estimation data using different ages as a predictor variable; this indicates which base age is best for predicting heights at other ages. The results are displayed in Fig. 4; the lowest root mean squared error is for a base age of 45, although the curve is very flat from 40 to 50 years and thus any base age within this range would be acceptable. Although, clearly, young base ages are not appropriate, we suggest that our results are largely dependent on the range of ages in the data; ages that are intermediate would be best simply because extreme ages would poorly predict height at the opposite extreme of age. Ideally, selection of a base age should incorporate (*i*) distribution of site quality for the population; (*ii*) ranges of ages to which the equations would be applied, for each site index; (*iii*) yield for stands of different age and site index. The selection procedure should be devised such that error in prediction of volume is minimized. As we lack the necessary information, we are forced to assume that the estimation data are representative of the population to which the equations are to be applied and conclude that a base age of 40 to 50 years is appropriate.

Conclusions

We found that parameter estimates were insensitive to the error structures we tried. However, the standard errors of the parameters were somewhat sensitive, particularly when comparison was between an assumption of independence and any alternative autocorrelation structure. Thus, error structure could have a large impact when model selection includes testing significance of parameters.

Weighting is necessary in our difference equation as the variance is dependent on the difference between the observed ages. An analytic weighting function worked well for our model.

Model selection should proceed from parameter estimation for individual plots. Some variations of Chapman–Richards function for site index equations specify that one or more parameters are constants independent of site quality. Rather than test alternative, previously published models, emphasis should be placed on building a model that reflects the given data set.

The parameters of a difference equation will depend on which differences are used to fit the equation (previous measurement vs. all possible differences). We did not obtain suitable estimates when only the previous measurement was used to predict height. A simulation study is required to determine how this result is dependent on data and model. Potentially, a discrepancy could be due to the model being incorrect; ideally the parameter estimates should be consistent regardless of the structure of the data set. The discrepancy could be regarded as a sort of goodness of fit; however, more study is required to determine the behavior of this discrepancy.

Height-growth models often predict height at some age based on height at some other age (i.e., site index). However standard procedures assume that predictor variables are constant. By fitting base-age invariant site index equations, site index and height-prediction equations are fit simultaneously. Although it is known that individual equations will have lower variance (Curtis *et al.* 1974), the curves will be biased; that is, neither the height-prediction equation nor the site index prediction equation will possess a shape that represents the true functional relationship between height and age across levels of site index. For example, when predicting height based on site index, the predicted height at base age will be overestimated for low site index and underestimated for high site index. This behavior may be avoided by fitting difference equations. "Regression towards the mean" is also evident in the shape of the curves; there is typically some consistent pattern in the shape of the curves as site index increases, and this pattern is underestimated when either height or site index is represented as a function of the other.

Choice of base age is typically specified so that it is less than rotation age. We found that a base age of 40 to 50 was superior for predicting height at other ages; this result is likely to be highly dependent on the range of age in the data set. Ideally, base age should be chosen such that variance of the volume estimates for the forest of interest are minimized; this requires that site index equations be integrated into a growth and yield system.

Acknowledgements

W.H. Carmean and D.J. Lenthall provided the data set for this paper. Dr. Carmean also reviewed an earlier draft. R.A. Monserud and an anonymous reviewer provided comments that helped to clarify various points. Partial funding for this work was provided by a grant to S. Zedaker, Virginia Polytechnic and State University.

Beck, D.E. 1971. Height-growth patterns and site index of white pine in the southern Appalachians. For. Sci. **17**: 252–260.

Bertalanffy, L. von. 1941. Stoffwechseltypen und Wachstumstypen. Biol. Zentralbl. **61**: 510–532.

Biging, G.S. 1985. Improved estimates of site index curves using a varying-parameter model. For. Sci. **31**: 248–259.

Borders, B.E., Bailey, R.L., and Clutter, M.L. 1988. Forest growth models: parameter estimation using real growth series. *In* Forest Growth Modelling and Prediction. Proceedings of the IUFRO Conference. *Edited by* A.R. Ek, S.R. Shifley, and T.E. Burk. USDA For. Serv. Gen. Tech. Rep. NC-120. pp. 660–667.

Brickell, J.E. 1966. Site index curves for Englemann spruce in the northern and central Rocky Mountains. USDA For. Serv. Res. Note. INT-42.

Bull, H. 1931. The use of polymorphic curves in determining site quality in young red pine plantations. J. Agric. Res. **43**: 1–28.

Burkhart, H.E., Tennent, R.B. 1977. Site index equations for radiata pine in New Zealand. N.Z. J. For. Sci. **7**: 408–416.

Carmean, W.H. 1972. Site index curves for upland oaks in the Central States. For. Sci. **18**: 109–120.

Carmean, W.H. 1975. Forest site quality evaluation in the United States. Adv. Agron. **27**: 209–267.

Carmean, W.H., and Lenthall, D.J. 1989. Height-growth and site-index curves for jack pine in north central Ontario. Can. J. For. Res. **19**: 215–224.

Carmean, W.H., Hahn, J.T., and Jacobs, R.D. 1989. Site index curves for forest tree species in the eastern United States. USDA For. Serv. Gen. Tech. Rep. NC-128.

Chapman, D.G. 1961. Statistical problems in population dynamics. *In* Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley. pp. 153–168.

Cieszewski, C.J., and Bella, I.E. 1989. Polymorphic height and site index curves for lodgepole pine in Alberta. Can. J. For. Res. **19**: 1151–1160.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., *et al.* 1983. Timber management: a quantitative approach. John Wiley & Sons, New York.

Curtis, R.O., DeMars, D.J., and Herman, F.R. 1974. Which dependent variable in site index – height – age regressions. For. Sci. **20**: 74–87.

Devan, J.S., and Burkhart, H.E. 1982. Polymorphic site index equations for loblolly pine based on a segmented polynomical differential model. For. Sci. **28**: 544–555.

Dyer, M.E., and Bailey, R.L. 1987. A test of six methods for estimating true heights from stem analysis data. For. Sci. **33**: 3–13.

Ek, A.R. 1971. A formula for white spruce site index curves. University of Wisconsin, Department of Forestry, Madison. For. Res. Note 161.

Furnival, G.M., Gregoire, T.G., and Valentine, H.T. 1990. An analysis of three methods for fitting site-index curves. For. Sci. **36**: 464–469.

Graney, D.L., and Burkhart, H.E. 1973. Polymorphic site index curves for shortleaf pine in the Ouachita Mountains. U.S. For. Serv. South. For. Exp. Stn. Res. Pap. SO-85.

Hahn, J.T., and Carmean, W.H. 1982. Lake States site index curves formulated. USDA For. Serv. Gen. Tech. Rep. NC-88.

Hamlin, D., and Leary, R.A. 1987. Methods for using an integro-differential equation as a model of tree height growth. Can. J. For. Res. **17**: 353–356.

Lenthall, D.J. 1986. Height growth and site index curves for jack pine (*Pinus banksiana* Lamb.) in the Thunder Bay area—a system of site quality evaluation. M.Sc.F. thesis, Lakehead University, School of Forestry, Thunder Bay, Ont.

Lundgren, A.L., and Dolid, W.L. 1970. Biological growth functions describe published site index curves for Lake States timber species. USDA For. Serv. Res. Pap. NC-36.

Monserud, R.A. 1984. Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type. For. Sci. **30**: 943–965.

Newberry, J.D. 1991. A note on Carmean's estimate of height from stem analysis data. For. Sci. **37**: 368–369.

Newnham, R.M. 1988. A modification of the Ek–Payandeh nonlinear regression moel for site index curves. Can. J. For. Res. **18**: 115–120.

Northway, S.M. 1985. Fitting site index equations and other self-referencing functions. For. Sci. **31**: 233–235.

Payandeh, B. 1974*a*. Formulated site index curves for major timber species in Ontario. For. Sci. **20**: 143–144.

Payandeh, B. 1974*b*. Nonlinear site index equations for several Canadian timber species. For. Chron. **47**: 194–196.

Pienaar, L.V., and Turnbull, K.J. 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.

Richards, F.J. 1959. A flexible growth function for empirical use. J. Exp. Bot. **10**: 290–300.

Ricker, W.E. 1973. Linear regressions in fishery research. J. Fish. Res. Board Can. **30**: 409–434.

Trousdell, K.B., Beck, D.E., and Lloyd, F.T. 1974. Site index for loblolly pine in the Atlantic coastal plain of the Carolinas and Virginia. USDA For. Serv. Res. Pap. SE-115.

Walters, D.K., Gregoire, T.G., and Burkhart, H.E. 1989. Consistent estimation of site index curves fitted to temporary plot data. Biometrics, **45**: 23–33.

Wittink, D.R. 1988. The application of regression analysis. Allyn and Bacon, Inc., Boston.

Wonnacott, R.J., and Wonnacott, T.H. 1976. Econometrics. 2nd ed. John Wiley & Sons, New York.