PRACTICAL METHODS FOR ESTIMATING NON-BIASED PARAMETERS IN SELF-REFERENCING GROWTH AND YIELD MODELS

Plantation Management Research Cooperative

Daniel B. Warnell School of Forest Resources
University of Georgia
Athens, Georgia 30602

PMRC TECHNICAL REPORT 2000-7

February 22, 2000

by

CHRIS J. CIESZEWSKI, MIKE HARRISON, AND STACEY W. MARTIN

PRACTICAL METHODS FOR ESTIMATING NON-BIASED PARAMETERS IN SELF-REFERENCING GROWTH AND YIELD MODELS

CHRIS J. CIESZEWSKI, MIKE HARRISON, STACEY W. MARTIN

D. Warnell School of Forest Resources
University of Georgia
Athens, GA 30602, USA

ABSTRACT. Since its inception in 1976, the primary objective of the Plantation Management Research Cooperative (PMRC) at the Warnell School of Forest Resources has been the development of growth and yield models for southern pines. Recent advances in computer and software technology have enabled the PMRC to explore the most computationally intensive data analysis methods. So-called traditional estimation techniques for dominant height projection equations involve the arbitrary choice of observed growth intervals and base ages, where each combination results in a different set of parameter estimates. Bailey and Clutter (1974) introduced the concept of base-age-invariant site index equations. In this paper, we present two practical approaches to this estimation technique and compare their results to those obtained from traditional methodologies. The base-age-invariant techniques produced identical results regardless of the choice of base ages or applied algorithm. The traditional methods were notably affected by the choice of base age and measurement intervals.

KEYWORDS: Base-age invariant, unbiased parameter estimates, nonlinear regression, dummy parameters.

1 INTRODUCTION

The average height of dominant and co-dominant trees at a given age is a critical component of growth and yield models for southern pine plantations. This statistic is little affected by the stand densities that are normally encountered in managed plantations. Thus, site quality estimation procedures based on stand height data are the most commonly used techniques for evaluating site productivity (Clutter, *et.* al, 1983). Most of these height-based techniques for evaluation of site quality rely on the development of site-index-based height over age curves hereafter called site index curves. Each site index curve defines an expected height/age relationship referenced by the expected height at a specified base age. Parameter estimation for site models involves many different statistical considerations such as criteria of fitting, error structures and independence of errors, homogeneity of variance and balance of data panels. We do not discuss here any of these considerations. We focus only on the dependence of model predictions on the use of base ages in model fitting.

In general, the most desirable data for the development of site index curves come from remeasured permanent sample plots or from stem analysis. Both of these data sources provide a number of observed height/age pairs for a given location and allow the ultimate flexibility in model forms and estimation techniques. Remeasurement or stem analysis data are usually combined by plots, resulting in an average height/age relationship for a given location. In order to reference the resulting curves by the height at a given age (site index), a choice of base age must then be made. This may suggest that site index be included in the model prior to its parameter estimation. For this reason a common approach to the parameter estimation in site index models is based on the assumption that the site indices are equal to the observed height values at the base age. This results in so-called base-age-specific parameter estimation using directly observed heights at specified base ages to restrict the predictions at these base ages to values equal to the observed data or deterministic data transforms. Most of the site index models in the literature are fitted this way. Since the observed height values contain measurement and sampling errors, such practice results in parameter estimates that are biased (Goelz and Burk 1996) and unique to the pre-selected base age (Heger 1973). This problem will persist even if multiple base ages are used in the parameter estimation process.

Exceptions are base-age-invariant methods of parameter estimation, some of which may also be biased. One example of a potentially biased base-age-invariant method is the guide curve system. The proportional guide curve system consists of constructing a single curve representing the mean heights at each age within the data. This guide curve is then proportionally scaled to pass through the observed reference height at any base age. The non-proportional guide curve system (Osborne and Schumacher 1935) modified the above guide curve method to scale the guide curves according to smoothed standard deviations of heights represented by the data at different base ages. Yet the most effective system for parameter estimation needs to be based on identifying individual trends represented in the data.

Having identified all the above problems of parameter dependence on selected base ages and the need for identification of individual trends, Bailey and Clutter (1974) suggested an estimation technique based on a simultaneous estimation of the common model parameters with the parameters that are site-specific. The site-specific parameters are equivalent to site indices and such an estimation technique is unaffected by any arbitrary choices of base ages, that is, it is base-age-invariant, and is based on identification consistencies of individual trends in the data. Bailey and Clutter (1974) used anamorphic and polymorphic variations of the following general model:

$$\log (H) = a + b(1/A)^c$$

where: a and b are regression parameters, either of which could be specific to the i^{th} site ($i = 1, 2, \ldots,$ m, where m is the number of represented sites) in which case they would be denoted as either a_i or b_i, and

$c > 0$ is an optional nonlinear parameter.

For example, the anamorphic model based on the above general equation is:

$$\log (H) = a_i + b(1/A)^c$$

The traditional methods for fitting site index equations involve the definition of the site-specific parameter, a_i, in terms of the height at a given base age (site index). The estimate of the site-specific parameter, therefore, is based on a single height observation for a given plot. This observation is subject to the measurement error associated with that data point, which in turn biases the fitting process because each curve is forced to pass through an erroneous point of the measurement error, which in turn biases the shape of the curve.

To avoid this problem, one can follow the procedure suggested by Bailey and Clutter (1974) and actually estimate the site-specific parameters, or the site index parameters, for all of the height/age data for each plot simultaneously while estimating the global model parameters. Parameter estimates resulting from such a procedure are unbiased and stable regardless of the choice of base age.

The purpose of this report is to describe two methods for fitting nonlinear, unbiased site index equations using PC-SAS. Both approaches were evaluated using three different choices of base age. The resulting models were compared to site index models fitted using the traditional biased techniques.

2 DATA AND METHODS

The data for this study came from a series of loblolly pine growth and yield plots established at six locations across the state of Georgia. The main objective of the study was to provide real growth series data for loblolly pine plantations managed under various silvicultural treatment regimes. Study sites were located near weather monitoring stations so that the effects of climate and atmospheric pollution factors on pine growth could be evaluated and modeled where appropriate (Borders and Bailey, 1997). The original study plan called for two complete blocks, each containing four 3/8-acre treatment plots. One of the following treatments was applied to each plot:

- H – Use herbicides to control all competing vegetation throughout the life of the study.
- F – Fertilize as follows: First two growing seasons, 250 lbs/ac DAP + 100 lbs/ac KCL in the spring and 50 lbs/ac of ammonium nitrate in mid summer. During each subsequent growing season, 50 lbs/ac of ammonium nitrate were applied in early to mid summer.
- HF – Both H and F treatments.
- C – Control treatment, not treatment following initial site preparation.

In addition to the above, the treatment plots at each location were replicated at various points in time. The first series of plots was established near Waycross, GA in 1987. Additional installations were established at the same location in 1989 and 1993. All plots have been measured annually since the first growing season.

To develop the methodology for unbiased estimation of site index model parameters, a single treatment was chosen from the four described above. This eliminated the need to model any treatment effects in addition to the height/age relationships. The vegetation control treatment (H) was chosen for this exercise. This resulted in 26 plots and 232 total observations.

The Chapman-Richards equation has been successfully applied by others to model height/age relationships for southern pines (e.g., Newberry and Pienaar, 1978; Pienaar and Shiver, 1980). Two forms of this model were used in this study:

Variable base age:

$$Hd_2 = Hd_1 \left[\frac{1 - \exp(-aA_2)}{1 - \exp(-aA_1)} \right]^b \tag{1}$$

and Fixed base age:

$$Hd = S \left[\frac{1 - \exp(-aA)}{1 - \exp(-aA_b)} \right]^b \tag{2}$$

where Hd_i = average dominant/codominant height at time i,
A_i = age at time i,
S = site index, defined as Hd at any given base age A_b,
a,b = parameters estimated from data.

3 MODELING APPROACH I: CONSECUTIVE REITERATIONS

The first approach to fitting the base-age-invariant site index equation uses nested iterative procedures to fit the global and site-specific parameters and it is due to Tait *et.* al. (1987). The procedure begins by fitting the global parameters (a and b) in equation (2), using the observed site index (S) for each plot. The site index in this case is defined as the average dominant/codominant height at any given base age. In the second step, estimates of the global parameters are used as constants and the site-specific parameter (S) is estimated for each plot. The observed S values for each plot are used as starting values for the fitting procedure. Next, the estimated S values for each plot become the "observed" values and the global parameters are refit. This procedure is repeated until the global parameters stabilize. Figure 1 illustrates with a flowchart how this iterative procedure was designed and programmed in SAS.

The initial data setup to implement the iterative procedure is straightforward. First, the base age must be chosen and the average dominant/codominant height for a given plot at the base age is defined as the site index (S_i). The site index value must be placed on each observation for each plot. The rest of the data consists of a plot identifier, age and the average dominant/codominant height at that age. Table 1 shows an example data set for the iterative procedure.

The iterative procedure was run on the 232 observations using site index values with base ages at 10, 5 and 3 years. For base age 10, the procedure stabilized after 8 iterations and took approximately 25 seconds to run on a Pentium Pro, 233 Mhz processor. For a base age of 3 years, the procedure stabilized after 210 iterations and took approximately 9 minutes. The convergence criteria in PROC MODEL (FIT statement) was then changed to 0.0000001. This setting increases the run time, but helps avoid converging to a local minimum.

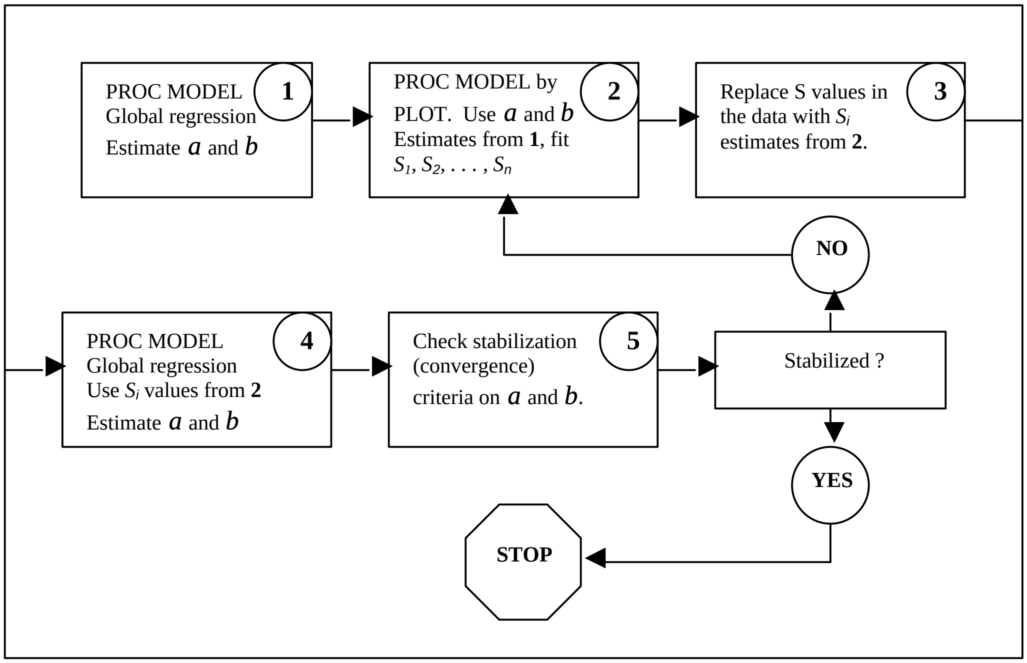

Figure 1. Iterative procedure to fit a base-age-invariant site index equation using PC-SAS.

Table 1. Example data setup for the iterative fitting procedure.

Plot	Age	Hd	S
1	3	14	49
1	4	18	49
1	5	24	49
:	:	:	:
1	10	49	49
2	3	12	46
2	4	17	46
2	5	22	46
:	:	:	:

4 MODELING APPROACH II: DUMMY VARIABLE APPROACH

The second approach to fitting a base-age-invariant site index equation is a dummy variable procedure that runs in a single PROC MODEL (SAS Institute, 1993) step. The following model was fitted:

$$Hd = (S_1 p_1 + S_2 p_2 + \ldots + S_n p_n)\left[\frac{1-\exp(-aA)}{1-\exp(-aA_b)}\right]^b \tag{3}$$

where S_i, = site-specific parameter for plot i,
p_i = dummy variable; = 1 for plot i; = 0 otherwise,
A_b = site index base age,
all others previously defined.

The first sum of terms, containing site-specific parameters and dummy variables , collapses for each plot into a single parameter unique to each plot during the fitting process. The observed site index values at the specified base age are used as starting values for the S_i parameters for each plot. The starting values are loaded via the ESTDATA data set in PROC MODEL. This data set consists of a single observation containing the starting values for each parameter. Table 2 shows an example of the input data setup for the dummy variable procedure. Table 3 illustrates the ESTDATA data set.

The necessary SAS code to complete this fitting procedure is presented in the appendix. The dummy variable approach run on the same 232 observations took approximately 9 seconds to converge to the final parameter estimates using the convergence criteria of 0.0000001. The low tolerance convergence criteria ensures consistency between the two fitting procedures by avoiding converging to a local minimum.

5 RESULTS

Using both the iterative and dummy variable procedures with base age site indexes at 3, 5 and 10 years, the results were identical in all cases. The following model was obtained:

Table 2. Example data setup for the dummy variable fitting procedure.

Plot	Age	Hd	p_1	p_2	p_3 . . .	p_n
1	3	14	1	0	0	0
1	4	18	1	0	0	0
1	5	24	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	10	49	1	0	0	0
2	3	12	0	1	0	0
2	4	17	0	1	0	0
2	5	22	0	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3. Example setup for the ESTDATA dataset for the dummy variable fitting procedure.

a	*b*	S_1	S_2	S_3 . . .	S_n
0.07	1.4	49	46	52 . . .	44

$$Hd = S\left[\frac{1-\exp(-0.090224A)}{1-\exp(-0.090224A_b)}\right]^{1.42671} \qquad \text{MSE} = 2.73;\ R^2 = 0.9826 \tag{4}$$

Figure 2 shows the height growth curve implied by equation (4) with a base age 10 and site index of 45.47 feet. The observed average dominant/co-dominant heights are also shown.

For comparison purposes, equation (2) was fitted to the same 232 observations using the ordinary, nonlinear least squares technique in PROC MODEL. The following models were obtained with site indices defined at base ages 3, 5 and 10 years:

Base age 10

$$Hd = S\left[\frac{1-\exp(-0.091793A)}{1-\exp(-0.091793A_b)}\right]^{1.43346} \qquad \text{MSE} = 3.57;\ R^2 = 0.9772 \tag{5}$$

Base age 5

$$Hd = S\left[\frac{1-\exp(-0.088226A)}{1-\exp(-0.088226A_b)}\right]^{1.349} \qquad \text{MSE} = 10.55;\ R^2 = 0.9327 \tag{6}$$

Figure 3 shows the height growth curves implied by equations (4),(5),(6) and (7) with a base age of 10 and a site index of 46 feet.

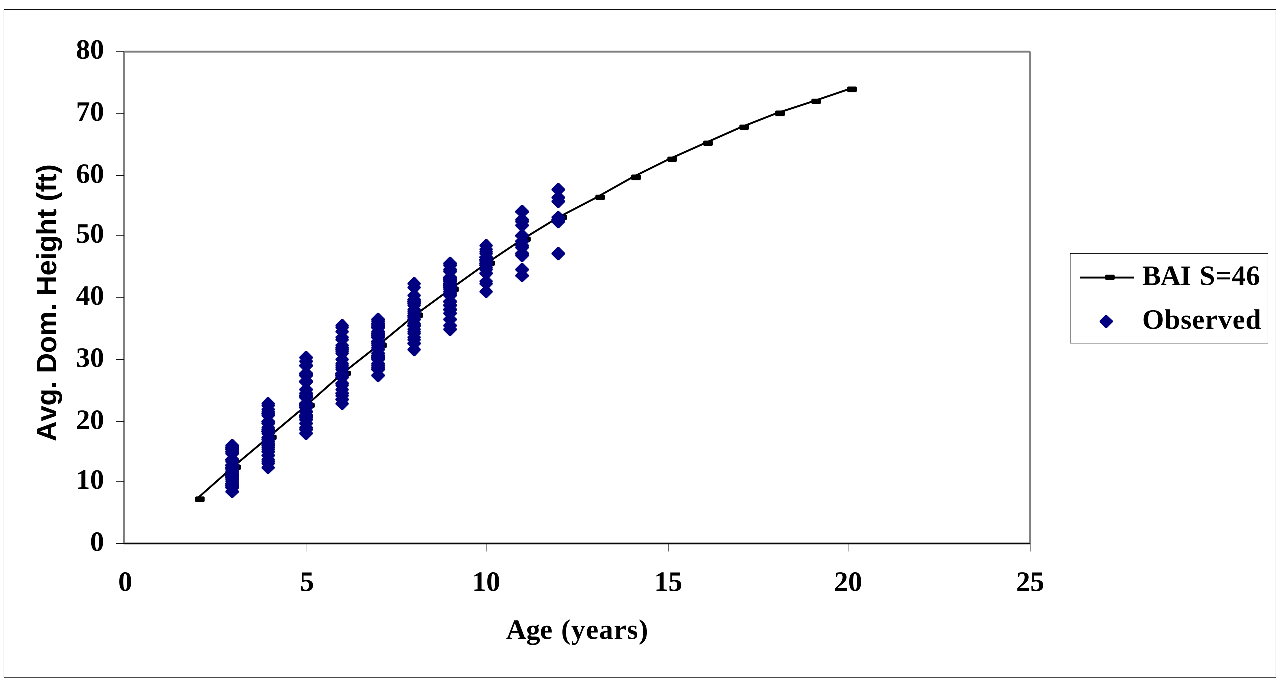

Figure 2. Predicted average dominant/co-dominant height growth curve implied by equation (4) and observed height measurements.

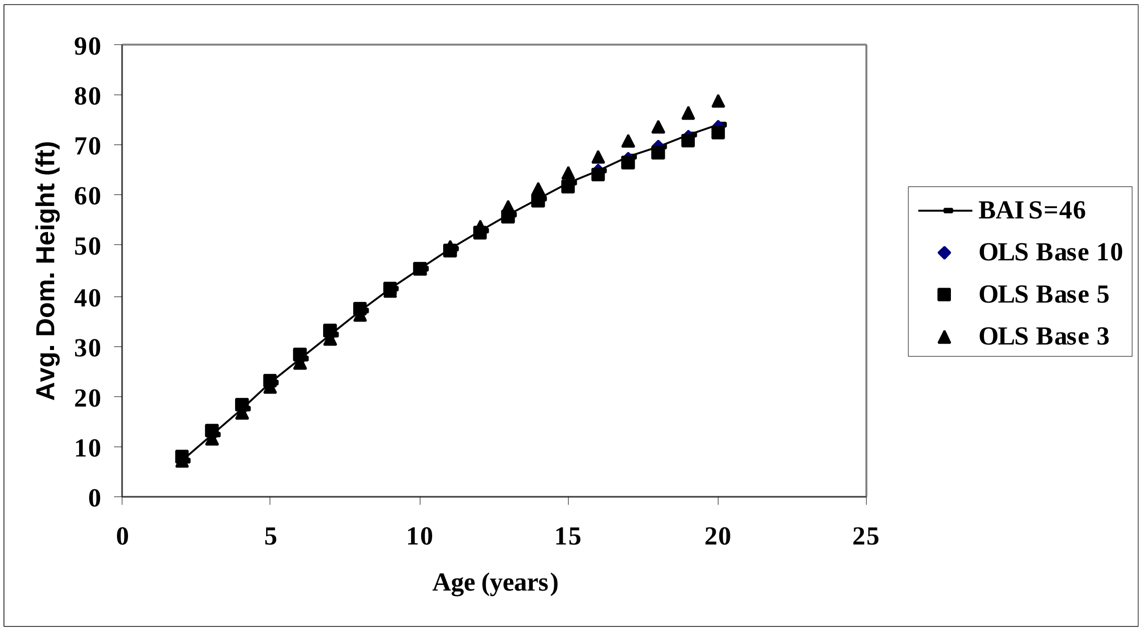

Figure 3. Predicted average dominant/co-dominant height growth curves for models fit with the base-age-invariant technique and the ordinary least squares technique for three different base ages.

Base age 3

$$Hd = S\left[\frac{1-\exp(-0.072765A)}{1-\exp(-0.072765A_b)}\right]^{1.40001} \quad \text{MSE} = 24.95;\ R^2 = 0.8400 \qquad (7)$$

6 CONCLUSIONS

Using fairly straightforward programming with the PROC MODEL and the SAS Macro Language (SAS Institute, 1990), the base-age-invariant (Bailey and Clutter 1974) site index equation and other similar growth and yield models can be fit using sufficient data without violating regression assumptions on error-free independent variables. The data must consist of repeated measurements on some multiple plots. The number of measurements per each plot has to be at least two, and the number of plots must be at least the number of global parameters considered in a model + 2. Unlike the traditional approach, where the heights at base age are not used to produce residuals in model fitting because they are used as site indices, with the methods described here all of the data points are used to produce residuals. There is no need to make any arbitrary choice regarding measurement intervals. In this exercise we chose plots that received the same silvicultural treatment regime in order to avoid the need to model treatment response in addition to the basic height growth pattern.

Compared to site index equations fit with ordinary, nonlinear least squares, the base-age-invariant model is more consistent. The traditional technique requires an arbitrary choice of base age prior to fitting the model. The model is then forced through the chosen height/age point. In the base-age-invariant technique, we recognize that each measurement is made with error and, therefore, it seems unreasonable to force the model through any given measurement. Instead, the curve is fit to the observed trend in the data.

The described base-age-invariant approaches are similar to mixed-effect modeling. The so-called local, or site-specific, parameters are in fact the random effects with a non-zero mean equal to the fixed effect representing average reference height at the given base age. In other words, they are sums of the random effects and the particular fixed effect corresponding to the site-specific parameter (while this fixed effect is not modeled explicitly, it can be easily recovered after the model fitting). All the other fixed effects are simply the global model parameters.

The main differences between the proposed methods and the formal mixed effects modeling are:

i) The estimation of the random effects is performed with the transformed model in which the reference height (sum of the random stochastic element and the mean height) becomes so called "expected-value parameter". This makes the parameter closer to linear improving its estimation properties and it should be considered as an improvement.
ii) The random effects are estimated on the expected-value-parameter transformed model with improved distributional properties of the random effects.
iii) The global criterion for estimation of the fixed and random effects is least square rather than the maximum likelihood, which might be considered a shortcoming.
iv) Both the random and the fixed effects are estimated simultaneously, which is an advantage.

This work is preliminary in that the choice of data may not have been ideal. We would like to evaluate the method on data with older ages and more inherent site variation. We would also like to evaluate the fitting technique with other dependent variables such as per-acre basal area, survival and per-acre yield. The ultimate application may involve using the technique to fit a system of seemingly unrelated equations with more than one site-specific parameter.

REFERENCES

Bailey, R.L. and J.L. Clutter. 1974. Base-age invariant polymorphic site curves. For. Sci. 20(2): 155-159.

Borders, B.E. and R.L. Bailey. 1997. Loblolly pine - pushing the limits of growth. Univ. of Ga. Consortium on Accelerated Pine Production Studies. Tech. Rep. CAPPS 1997-1.

DuPlat, P. and M. Tran-Ha. 1986. Modeles de Croissance en Hauteur Dominante pour le hêtre, le sapin pectiné et le pin sylvestre (dans le Massif de l'Aigoual). Document No. 86.1. Office National de For ê ts. France. (In French).

Goelz, J.C.G., and T.E. Burk. 1996. Measurement error causes bias in site index equations. Can. J. For. Res. 26: 1586-1593.

Heger, L. 1973. Effect of index age on the precision of site index. Can. J. For. Res.. 3:1-6.

Newberry, J.D. and L.V. Pienaar. 1978. Dominant height growth models and site index curves for site-prepared slash pine plantations in the lower coastal plain of Georgia and North Florida. Univ. of GA. Plantation Man. Res. Coop. Res. Paper No. 4.

Osborne, J.G. and F.X. Schumacher. 1935. The construction of normal-yield and stand tables for even-aged timber stands. J. Agric.\ Res.\ 51:547---564.

Pienaar, L.V. and Shiver, B.D., 1980. Dominant height growth and site index curves for loblolly pine plantations in the Carolina flatwoods. So. J. Appl. For. 4: 55-59.

SAS Institute, 1990. *SAS Guide to Macro Processing, Version 6, Second Edition*. SAS Institute, Inc. Cary, NC.

SAS Institute, 1993. *SAS/ETS User's Guide, Version 6, Second Edition*. SAS Institute, Inc. Cary, NC.

Tait, D.E., Cieszewski, C.J., and I.E. Bella. 1988. The Stand Dynamics of Lodgepole Pine. Can. J. For. Res. 18: 1255--1260.

APPENDIX A

```
%let mini=1;                          *Creates the macro variable mini equal to 1.;

data one;
infile 'c:\data.dta';                 *Inputs the data set data.dta where Hd=dominant height,;
input plot age Hd Ai Hdi;             * Ai=base age, and Hdi=dominant height at the base age.;

proc sort; by plot age;

data two;  set one; by plot age;
if last.plot and last.age then do;
call symput('maxi', trim(left(plot)));*Creates the macro variable maxi equal to the;
end;                                  * total number of plots.;

data three;  set two; by plot age;
array p(&maxi) p1-p&maxi;             *This data step creates an array of dummy;
array S(&maxi) S1-S&maxi;             *variables (p1, p2, …, pn) as shown in;
retain S1-S&maxi;                     *Table 2 and creates a horizontal vector;
p(plot)=1;                            *of observed site index values (S1, S2, …, Sn),;
do i=1 to &maxi;                      *as shown in Table 3, at the specified base age.;
if p(i)=. then p(i)=0;  end;  S(plot)=Hdi;

proc sort; by descending plot;

data start;
set three; by descending plot;        *This data step creates the temporary SAS data set start;
keep S1-S&maxi alpha beta;            *that contains the starting values for each parameter.;
if _n_ =1 ;  alpha=.8 /*starting value for alpha*/;  beta=1.4 /*starting value for beta*/;

%macro fits;                          *This is the SAS macro fits that contains;
proc model data=three;                * the model fitting procedure.;
exogenous age Ai %do i=&mini %to &maxi %by 1; p&i %end;;
endogenous Hd;  parms S1-S&maxi alpha beta;

Hd=(%do i=&mini %to %eval(&maxi-1)
 %by 1; p&i*S&i+ %end; p&maxi*S&maxi)*
((1-exp(-alpha*age))/(1-exp(-alpha*Ai)))**beta;

fit Hd / converge=.0000001 estdata=start;

run;
%mend fits;

%fits;

run;
quit;
```

The necessary SAS code for the dummy variable fitting approach is given above.

APPENDIX B

The following SAS code demonstrates an example of how to approximate unbiased parameter estimation by calculating average site index values predicted from all observations.

```
filename hts 'filename'
data AvrSI;
    infile hts firstobs=1;
    input PlNo TrNo t Ht MrmNo H1-h26;
    run;
proc nlin data=AvrSI;
        parms alpha=0.2 beta=100;
        jota=-1-alpha; delta=20*beta*50**jota; ksi=80*beta;
        array hx {26} h1-h26;  array x {26} x1-x26;
        do i=1 to MrmNo; x{i}=ksi*(i*10)**jota; end; SumSI=0;
        do i=1 to MrmNo; SumSI=SumSI+hx{i}+((hx{i}-delta)**2+hx{i})**.5; end;
        SI=(SumSI/MrmNo-delta)/2;
        model Ht=(SI+delta)/(1+20*beta*t**jota/SI);
        run;
```