United States
Department of
Agriculture

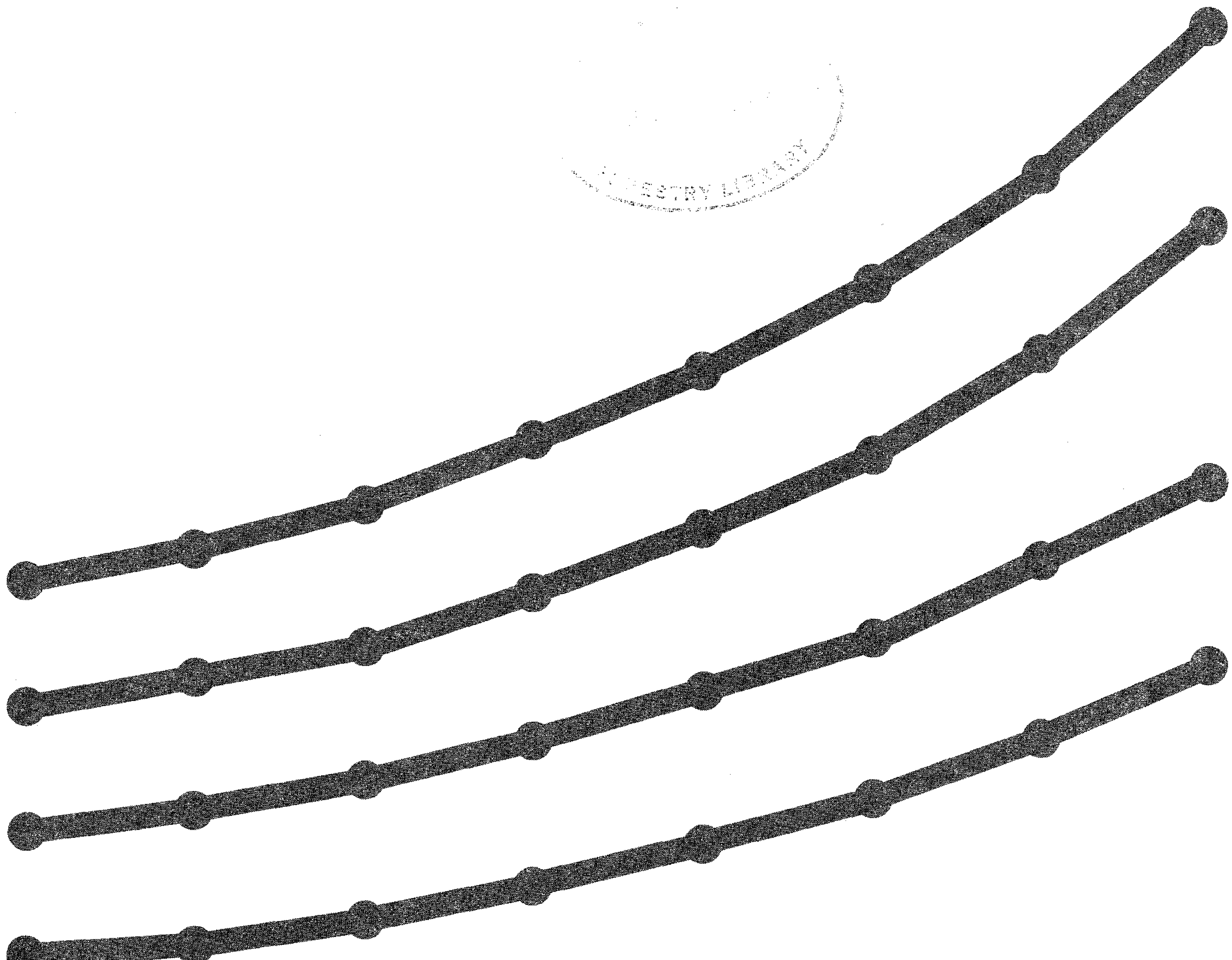
Forest
Service

North Central
Forest Experiment
Station

Research
Paper **NC-234**

An Evaluation of the STEMS Tree Growth Projection System

Margaret R. Holdaway and Gary J. Brand



CONTENTS

North Central Forest Experiment Station
Forest Service—U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, Minnesota 55108
Manuscript approved for publication January 31, 1983
September 1983

An Evaluation of the STEMS Tree Growth Projection System

Margaret R. Holdaway, *Mathematical Statistician*
and **Gary J. Brand,** *Research Forester*

In 1975, research was begun at the North Central Forest Experiment Station to develop a system to project tree growth in the Lake States region. The tree growth projection system developed (USDA Forest Service 1979a) represented just one component of a broader Forest Resources Evaluation Program (FREP). The original projection system (which we will call GTR49) has been modified several times; the current version is STEMS, Stand and Tree Evaluation and Modeling System (Belcher *et al.* 1982).

STEMS is an individual tree, distance-independent model that can be used to project diameter growth and mortality for any species mixture and stand structure. Growth and mortality coefficients have been developed for the major tree species in the Lake States region (Michigan, Minnesota, and Wisconsin). However, because the current version of STEMS contains an incomplete regeneration submodel, it should not be used to project forest stands where *seedlings* are an important component.

Before any model can be used effectively, it must be evaluated. The purpose of model evaluation is to increase one's "confidence in the predictive capabilities of the model" (Schaeffer 1980) or to compare "its agreement with the real-life system it is intended to represent" (Goodall 1972). In the model evaluation process the primary questions include, who is going to use the model and their purpose for using it. As a result, model evaluation is very subjective and there is no absolute test of validity.

Model evaluation should not attempt to determine whether or not a model is acceptable. Because models are used for many different purposes—e.g., studies on competition and succession change, updating inventories, studying the impact of different cutting prescriptions—acceptability will depend on the use. For this reason, a more appropriate approach to model evaluation is to present an extensive body of quantitative information on the model's performance. These results can provide the basis for users to evaluate the model for their own specific purpose.

The purposes of this paper are (1) to provide a brief preliminary analysis of how well the performance of STEMS compares with GTR49, and (2) to report the results of an analysis of the STEMS tree growth projection system for a wide range of forest conditions and identify its main strengths and weaknesses. These conditions encompass a wide range of forest types, species, time intervals, stand densities, and site qualities. Evaluations concentrate on three key variables—tree d.b.h. (evaluates the growth component of the model), number of trees per acre (evaluates the mortality component), and stand basal area (evaluates the entire model). The analysis was performed on a systematic sample from the calibration data base as well as on five independent data sources. The results are presented in a simple format that can be interpreted without specialized statistical knowledge. Belcher, Holdaway, and Brand (1982) have previously reported some preliminary results of STEMS' performance.

EVALUATION PRINCIPLES

Evaluation Philosophy

Modelers need not be concerned with proving the "truth" of a model because no model will behave exactly like the system it is modeling. All models are imperfect; at best they are a simplification of extremely complex biological and ecological processes (Goodall 1972, Goulding 1979, Pilgrim 1975, Taylor 1979, Valentine 1978). The objective, then, of the evaluation process is not to accept or reject the model as true or false but to determine the quality of the predictions (Goodall 1972). There is no absolute test of validity or accuracy of a model but only subjective judgments based on the proposed use of the model, the acceptable level of errors, the availability of alternative models, and other user-related practical considerations. Therefore, we will not be concerned with hypothesis tests but with quantitative statements about accuracy and precision.

Furthermore, models can be expected to fail on some occasions due to oversimplification of assumptions and relations, rarely encountered situations that have not been properly modeled, or extreme conditions that have not been considered. Therefore, it is unwise to base an evaluation of the model's performance solely on prediction errors averaged over all plots without regard to the more subtle discrepancies. A modeler needs to identify those areas in which the model produces fairly reliable results (called the range of applicability) and to highlight the weak areas (Goulding 1979, Pilgrim 1975, and Taylor 1979). One way to do this is to examine patterns of residual errors for various time intervals and various initial conditions (Pilgrim 1975, Taylor 1979). These patterns give much more information about the adequacy of the model than any single test of significance. It is also important to consider the model's performance outside the range of the calibration data and to determine whether the results are consistent with current knowledge. So modelers are not just concerned with how well the model functions, but also with the range of conditions over which it functions well.

Once the user has determined where and under what conditions the model performs adequately, he can judge the model's practical value. This may be a more desirable criterion of acceptance than a statistical measure or test that gives little information on exactly what is happening (Pilgrim 1975, Goodall 1972, Goulding 1979). Thus, the critical question is not whether the model is valid but whether it is useful (Mankin *et al.* 1975).

Therefore, recognizing the subjective, user-dependent nature of model evaluation, our approach is to provide a general overview along with a detailed analysis to assist the user in making that evaluation.

Calibration Data and Independent Data

We believe model evaluation should begin by examining the model's performance on the calibration data base. Any marked bias would indicate a basic flaw in the mathematical model used to describe the system. Leary *et al.* (1979) presented three problem areas that may cause this bias:

(1) the mathematical equations fail to describe parts of the process,
(2) model components are combined incorrectly (i.e., additive or multiplicative), and
(3) the numerical constants obtained from the calibration are in error.

If the model fails at this stage, the model will need to be refined before proceeding with further model evaluation.

A projection system that performs well on calibration data may still fail to satisfactorily fit independent subsets of the region. Therefore, the model should also be tested against independent data before its performance can be accurately assessed (Schaeffer 1980, Taylor 1979). This data should cover as wide a range of conditions as possible.

Comparison of Models

There is no universally accepted criterion against which the model's performance can be compared; validation is never absolute. Furthermore, we often compare the relative performance of two alternative models. This is how models are progressively developed and improved (Goodall 1972). The new model replaces the old if it is judged to perform better.

When comparing alternative models, it is unlikely that one model will give better predictions for all variables over a full range of stand conditions and forest types. For the *user*, the choice of a model will depend on which variables are more important, the amount of bias and variability in the predictions, the seriousness of the problem areas, and many other less obvious factors. The appropriate weighting of these various considerations depends on the purposes for which the model is to be used and is often very subjective and arbitrary (Goodall 1972). For the *model developer* other important considerations include the simplicity of the model, the ease of recalibrating the model, and the number of unknown coefficients to be estimated.

Data Base Differences

Differences between the calibration and validation data bases should be carefully considered when evaluating the model. For example, the types of plots used for the calibration may be different than those used for the validation or new conditions may occur that are outside the range of the calibration data. Problems can also arise if the geographical location of the calibration data is different from the area where the projection system is applied.

Can the yields observed on *experimental* plots be duplicated on general forest land? The tree growth model coefficents were developed mainly from research plots. Such plots usually are chosen for their good location and minimum natural damage and also have been carefully managed so the result is a more uniform stand than would be found on general forest land. Consequently, we would expect the model coefficients to overpredict production when applied to general forest land. Bruce (1977) has noted these growth differences but concludes that "no single magic number" can be used to adjust for it.

Another problem may arise if the model has some weak areas that have no effect in one predictive situation but that surface in another—the model may be valid for one "range of conditions" but invalid for another (Schaeffer 1980). For example, if the data used to calibrate the model did not include very small trees and the model was later used to project growth on these trees, the results may be seriously in error. This type of problem can be evaluated by comparing various initial conditions on the calibration data base and on subsequent validation data sets.

Not only may a model be invalid under a certain range of conditions but also for different geographical locations due largely to climate and soil factors.[1] How can these location differences be handled, especially when the model is applied outside the region in which it was calibrated? One factor in choosing a model form should be the ease in adjusting the coefficients so that they can be altered to account for local conditions. An adjustable model needs to be fairly easy to understand and should not require complex manipulations to change the amount of growth or the relative allocation of growth to trees on a plot.

EVALUATION ATTRIBUTES

Accuracy and Precision

The quality of the predictions made by a tree growth projection system can be analyzed in terms of both accuracy and precision. Predictions with small systematic errors are said to be highly accurate. A common measure of accuracy is the mean error. The mean error can be misleading, though, if there are large offsetting positive and negative random errors. Therefore, a measure of variability is also needed. Predictions with small random errors are considered to be highly precise and hence will be reproducible. The standard deviation (or its square, variance) is the most common of the many measures of precision. Our results will include both accuracy and precision using the mean and standard deviation of the prediction errors.

Evaluation Variables

Judgments about a tree growth model's performance should be based on key variables. We wanted to choose a set of variables that covered the most important aspects of stand and tree growth and that were directly estimated quantities. At the stand level, the key variables we use are stand basal area per acre (BA) and number of trees per acre (NT). Not only is BA an important stand attribute, but it is also an important component in estimating volume. NT is used to evaluate the mortality component, which is important because mortality greatly influences overall model performance. At the tree level, the variable we use is tree d.b.h. (DBH), which is important in management strategies, tree volumes, and mortality estimates. Modelers may find this detailed tree level analysis especially beneficial. Therefore, throughout this paper we will evaluate how well the model predicts BA and NT at the stand level and DBH at the tree level.

Components

When a model has been built from components describing a number of separate processes, it is informative to evaluate the model in a way that will show how well each process is working. In GTR49 and STEMS, DBH measures how well the growth component functions and NT measures how well the mortality component functions. BA, a combination of the two, measures how well the model functions as a whole. Errors in predicting BA can be due to problems in the growth component and/or mortality component. BA prediction could be very good even with large, but cancelling, errors in both DBH and NT. Therefore, it is the indepth evaluation of the separate growth (DBH) and mortality (NT) components that provide the clearest picture of how the system is operating.

A secondary component also needs to be tested. Whenever tree crown ratio is unknown, it is estimated using a crown ratio function. Errors resulting from this approximation are studied in an independent analysis of the crown ratio function.

Time Effect

Continuous growth models will predict nearly perfectly over very short time intervals. The longer the

[1]*This aspect of the validation process is highly complex and we will not discuss it in this paper.*

prediction interval, the less accurate the prediction will be (Goodall 1972). Consequently, if the time intervals used in the evaluation vary greatly, it may be difficult to make any justifiable conclusions. Two obvious choices can be made: (1) only use data from plots with similar time intervals or (2) standardize all results to a common time interval, such as 10 years. We have taken the second approach to make full use of the data. Then, we can compare plots with different measurement lengths or pool results together for a general overview. To standardize the time effect we assumed that error decay (the way error changes over time) is linear. We will evaluate this assumption in greater detail later in the paper.

Besides standardizing the time effect for most aspects of this study, we also analyzed the time effect in two ways. The first approach is an analysis of error decay on a property with at least three measurements. From this it is possible to analyze how accuracy and precision behave over time. The second approach is a long-term projection to test whether the growth model produces plausible results when extended far beyond the time frame in which it was calibrated.

All the results presented in this paper have been standardized to 10 years except those presented in the section on error decay and long-term projections.

METHODS AND PROCEDURES

Data Bases

The calibration data base consisted of tree measurements from permanent growth plots in the Lake States (fig. 1) (Christensen *et al.* 1979).[2] The plots ranged in size from 0.1 to 0.5 acre. On many of the plots only trees more than 5 inches d.b.h. were remeasured (but this diameter limit was as low as 0.6 inches on some plots and as high as 9 inches on others).

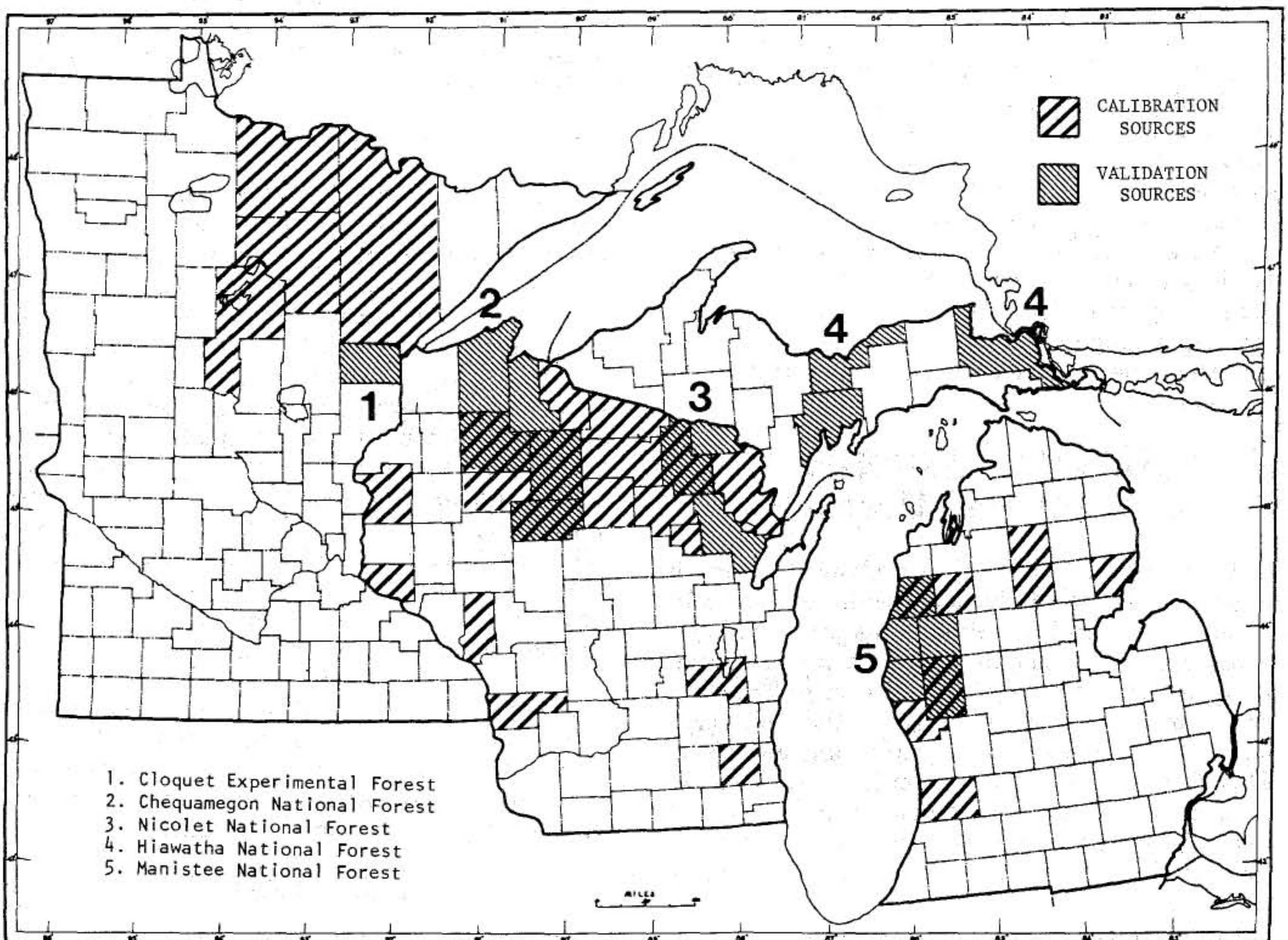


Figure 1.—*Geographical location of the calibration data and the five validation data sources used to test the projection system. Counties are shaded if they contain at least one plot.*

The data from the Lower Peninsula of Michigan are entirely red pine and jack pine plantations. The Wisconsin data are mostly northern white-cedar and natural hardwood stands (oak-hickory type in the south and west and northern hardwoods in the north). There is also a red pine plantation adjacent to the Upper Peninsula of Michigan and a white spruce plantation about 150 miles southeast of it. The northern Minnesota studies include natural and planted red pine and jack pine stands, and natural quaking aspen, white pine, and black spruce stands.

We systematically sampled every fifth calibration plot to yield 293 plots with approximately 7,700 live trees. Only those trees that were alive at the first measurement and were both predicted alive and observed alive at the last measurement were analyzed at the tree level. On the calibration sample, the average initial age was 48 years, the average initial tree diameter was 6.6 inches, and the average time from initial to final measurement was 20 years (fig. 2). The 80+ class for initial age contains observations up to 144 years. The validation data base came from five independent sources or properties in the Lake States (fig. 1):

(1) Cloquet Experimental Forest of the College of Forestry, University of Minnesota;[3]
(2) Chequamegon National Forest, Wisconsin;[4]
(3) Nicolet National Forest, Wisconsin;[4]
(4) Hiawatha National Forest, Michigan;[4]
(5) Manistee National Forest, Michigan.[4]

The Cloquet plots were all one-seventh acre with diameters of all trees more than 4.9 inches recorded. The National Forest plots were 10-point clusters of variable radius plots with measurements on trees 1.0 inch in diameter and larger. All National Forest plots were measured roughly 10 years after the initial measurement.

The total validation data base from the five properties contained 822 plots with nearly 11,200 trees. The average initial age of the plots was 46 years (only 2 years younger than the calibration data), the average initial diameter was 8.2 inches, and the average time interval was 13 years (much shorter than the calibration data) (fig. 2). Again the 80+ class for initial age contains observations up to 196 years.

Each plot in the two data bases contained site index, stand age, measurement dates, and a list of trees. Each tree in the list had a species code, tree factor (trees per acre represented by the tree), and crown ratio code[5], plus for each measurement a diameter and status code (alive, dead, or cut).

Basic Projection Data

We used the following steps to assemble the projection data.

(1) Project the growth from the years of the first measurement to the year of the last measurement, using the tree and stand characteristics at the first measurement as initial conditions for the projection.
(2) Remove a tree from the projection tree list in the same year it was observed to have been cut.
(3) Add a tree to the list of trees being projected in the same year it was observed to be an ingrowth tree.
(4) Write one output file of stand information containing initial conditions along with final predicted and observed conditions.
(5) Write a second output file of similar information for each tree carrying along pertinent stand information.
(6) Use these two files as the basic data for analyzing and summarizing the projection results at both the stand and tree levels.

The basic analysis involved calculating the error (i.e., predicted minus observed value) for a given attribute (DBH, NT, and BA) and then standardizing this error to 10 years using the equation:

$$\frac{\text{(predicted attribute}-\text{observed attribute)}}{\text{number of years in measurement interval}} \times 10.$$

[2]*We wish to thank the following persons for making information available to us and/or helping collect field measurements: Carl Tubbs, Dick Godman, Gus Erdman, Gib Mattson, John Benzie, Paul Laidly, and Bob Barse, North Central Forest Experiment Station; Rod Jacobs, USDA Forest Service, State & Private Forestry; Alan Ek, University of Minnesota, College of Forestry; Wisconsin Department of Natural Resources; John W. Moser, Purdue University; A. B. Johnson, Owens-Illinois Corporation; and Cal Stott, USDA Forest Service (retired).*

[3]*Dietmar Rose and the staff of the Cloquet Forestry Center made the past measurement data available and assisted in the 1976 remeasurement.*

[4]*The data for the four National Forests were provided by The Renewable Resources Evaluation Unit at the North Central Forest Experiment Station.*

[5]*Throughout this paper crown ratio is expressed as a code integer. A crown ratio falling between 1 and 10 percent was coded as 1, 11-20 percent was coded as 2,... and 91-100 percent was coded as 10.*

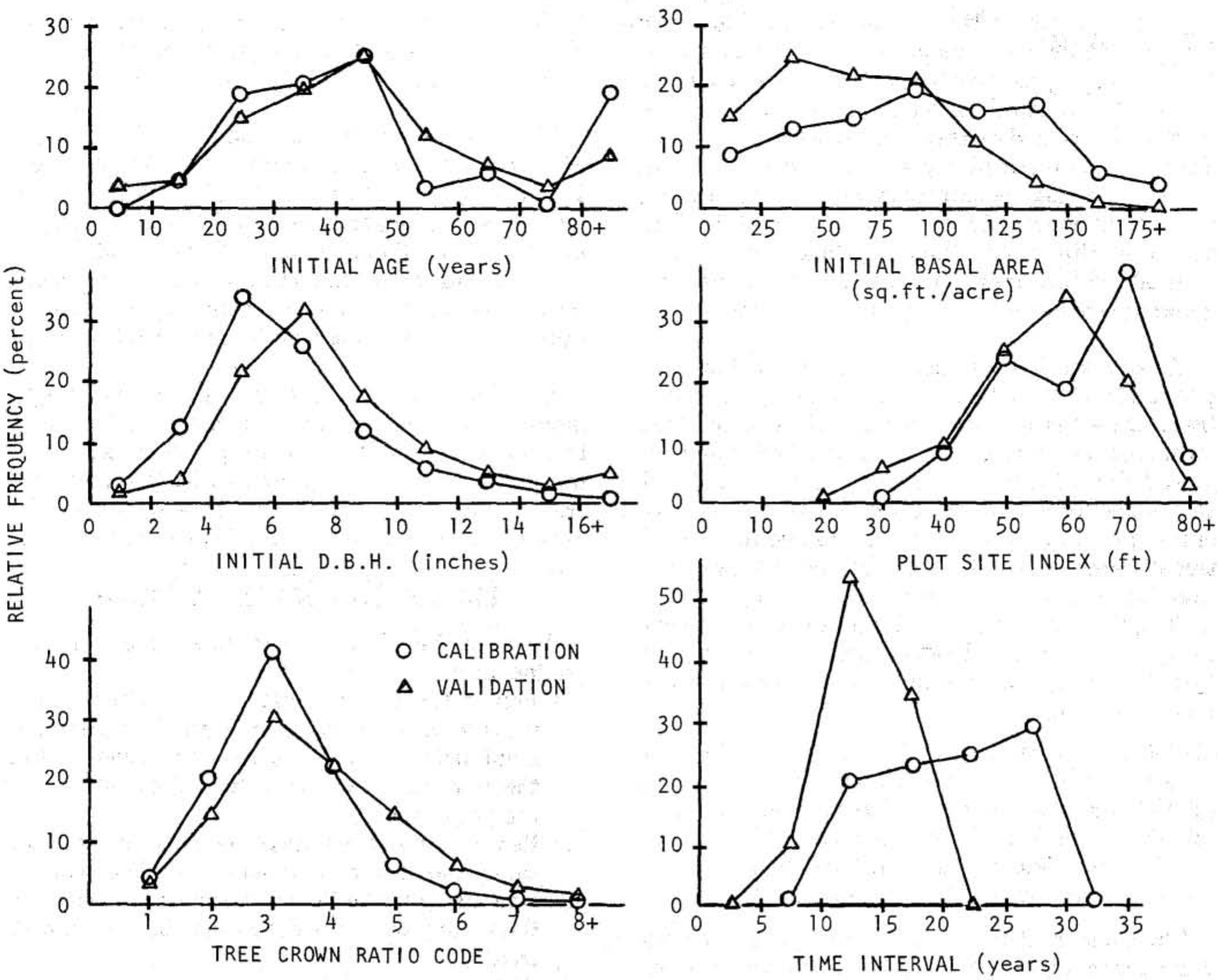


Figure 2.—*Distribution of 293 calibration plots and 822 validation plots in the Lake States by initial conditions for both stand and tree attributes.*

With this equation *an overprediction is a positive error and an underprediction is a negative error.* An accuracy and precision analysis was performed by calculating the mean and standard deviation of the 10 year errors over each class of interest such as forest type, species, property etc. If the predictions were perfect, all errors, and their mean and standard deviation, would be zero. The farther the mean and standard deviation are from zero, the greater the bias and the variability, respectively.

Model Differences

It is important to understand the basic differences in the two models before evaluating them. The original model, GTR49, is a *stand component* model. Each tree list is broken down into a maximum of four species groups and each species group into two size classes. These divisions are called components and hence the name "stand component" model. Each stand component is projected as a whole and later the component's growth is allocated to the individual trees. The current version, STEMS, contains an individual tree model that uses similar equations but deals with individual trees, not components of trees. Belcher *et al.* (1982) presents a detailed description of the model changes in STEMS, but the main differences between the two models are as follows:

(1) Growth potential—the same function and coefficients are used in both growth models but in STEMS they are applied to individual trees not stand components. (Both functions use site index and crown ratio as variables in this equation.)

(2) Site index—GTR49 used a single plot site index for all trees on the plot while in STEMS each species is given a site index. The conversion equations used to predict species site index from the plot site index were provided by Carmean (1979) and Carmean and Vasilevsky (1971).

(3) Crown ratio function—the GTR49 stand component approach has been replaced by an individual tree submodel in STEMS.[6]

(4) Modifier function—the GTR49 stand component submodel has been replaced by an individual tree modifier function[7] that includes an upper limit on the stand basal area.

(5) Interaction—GTR49 attempted to account for interaction between species classes in the modifier function, whereas STEMS does not.

(6) Allocation rule—this function in GTR49 has been removed, and its role is now handled by the modifier function in STEMS.

TEST RESULTS AND DISCUSSION

Model Comparison

In this section we will present a very general overview of the predictive ability of the two models using both the calibration and validation data bases (table 1).

[6]*The crown ratio function is reported in an unpublished manuscript by the senior author.*

[7]*The modifer function is reported in an unpublished manuscript by the senior author.*

A comparison of the performance of the two models using the calibration data shows both models doing well with little difference between them. However, the STEMS model produces lower standard deviations so it is slightly more precise. The performance of the two models on the calibration data base doesn't reveal any basic mathematical flaw in either model.

The performance of the two models for all the validation properties together shows similar positive mean DBH errors in 10 years. STEMS shows a small NT error whereas GTR49 underpredicts NT an average of 10 trees per acre in 10 years. This underestimation of NT combined with the positive DBH error results in a more accurate estimate of BA for GTR49 than for STEMS, even though STEMS is better on the growth and mortality components separately. In looking at the precision analysis, using the standard deviations, STEMS is slightly more precise than GTR49 for all three key variables—NT, DBH, and BA.

A more detailed analysis also showed that the models performed similarly, but we found three cases in which GTR49 did not predict as well as STEMS. GTR49 overestimated tree DBH's on stands with initial basal areas less than 25 square feet per acre, and it underestimated NT (overestimated mortality) on stands with more than 800 trees per acre or with basal areas of more than 150 square feet per acre. The two problems in the mortality component represent an increase in NT error of about 60 trees per acre, but these conditions only occur on 6 percent of the stands.

Table 1.—*Summary of the mean and standard deviation of the errors in 10 years for GTR49 and STEMS. The errors are evaluated for the attributes of tree d.b.h. (DBH), number of trees per acre (NT), and stand basal area per acre (BA). Errors are computed as predicted attribute minus observed attribute.*

Errors (10 years)	Calibration		Validation											
			All		Cloquet		Chequamegon		Nicolet		Hiawatha		Manistee	
	GTR49	STEMS	GTR49	STEMS	GTR49	STEMS	GTR49	STEMS	GTR49	STEMS	GTR49	STEMS	GTR49	STEMS
MEAN														
DBH (inches)	0.02	−0.03	0.13	0.11	0.15	0.06	0.04	0.13	0.06	0.09	0.15	0.09	0.24	0.28
NT (trees/acre)	4	6	−10	2	4	6	−17	16	−33	−18	−11	5	−6	2
BA (sq.ft./acre)	1.3	.6	1.9	3.5	4.7	4.2	−.0	4.7	−2.5	.8	.1	2.2	3.8	5.0
STANDARD DEVIATION														
DBH (inches)	.52	.46	.73	.63	.62	.54	.75	.63	.85	.69	.81	.67	.78	.69
NT (trees/acre)	36	37	81	78	20	20	105	108	103	104	116	100	72	66
BA (sq.ft./acre)	11.6	10.1	14.6	13.4	9.7	9.4	13.7	13.3	15.1	2.4	17.3	14.2	18.7	19.0
NUMBER OF PLOTS	293		822		292		123		145		114		148	
NUMBER OF TREES[1]	7,702		11,182		4,572		1,587		1,775		1,720		1,528	

[1]Values given are for the STEMS model.

One other major problem was found. A 100 year projection of red pine using GTR49 predicted basal areas of more than 450 square feet per acre with no apparent leveling off.

The two models differ in complexity. Schaeffer (1980) states "the simplest model which can be acceptably validated is deemed more suitable than the more complex model." Important in evaluation complexity is the ease in:

(1) calibrating or recalibrating the model,
(2) understanding the various components of the model,
(3) programming the model and understanding the projection program, and
(4) adjusting the model for specific user needs.

In all four of these areas we judged STEMS to be clearly preferable. GTR49 has some complex parts that were hard to calibrate and understand. Considering the factors of performance and model complexity, we recommend STEMS.

A more detailed description of STEMS' performance provides confidence and cautions in its application.

STEMS' Model Performance

Accuracy analysis

Growth component (DBH errors). We will first evaluate the growth component at a broad level by analyzing DBH errors by species for all the validation properties (table 2). Most species on the validation properties have a slight positive bias (overestimate growth). Those species with a mean DBH error of more than 0.30 inches in 10 years are: white spruce (0.32), northern white-cedar (0.37), basswood (0.38), white ash (0.35), and other red oak (0.76). However, all of these species have small mean DBH errors on the calibration data. Taking a weighted average of the validation and calibration totals shows only other red oak (at 0.74) with an error of more than 0.30. Northern white-cedar is the only species with consistently high values on all five validation properties. Diameter growth for all species combined shows a small but consistent overprediction on all five validation properties. In addition, diameter growth on one property, Manistee in lower Michigan, shows a clear overestimation for most species.

Mortality component (NT errors). Information on the mortality component is provided by the NT errors (table 3). Overestimating the number of live trees on a stand indicates that the mortality function is killing too few trees (underestimated mortality).

The four National Forests have variable radius plots with small trees (1 to 5 inch) measured on a 0.01 acre fixed radius plot. If even one of these small trees is assigned the wrong status by the mortality function, the estimated number of live trees per acre would be altered by 100 trees. This is the main reason why the errors were so much larger for these

Table 2.—*Summary of the mean d.b.h. errors (inches) by Lake States species in 10 years using the STEMS model. The number of tree observations appears in parenthesis. Positive values are overpredictions.*

Species	Calibration	Validation					
		All	Cloquet	Chequamegon	Nicolet	Hiawatha	Manistee
Jack pine	−0.01 (890)	0.13 (1,539)	0.06 (1,109)	0.15 (60)	0.08 (46)	0.40 (162)	0.34 (162)
Red pine	− .04 (3,221)	− .17 (1,360)	− .19 (871)	− .29 (64)	− .08 (69)	− .30 (134)	.00 (222)
White pine	− .11 (148)	− .01 (297)	− .19 (128)	− .02 (36)	− .11 (36)	.13 (73)	.65 (24)
White spruce	− .06 (156)	.32 (148)	− .02 (54)	1.12 (13)	.55 (49)	.22 (32)	—
Balsam fir	− .16 (402)	.11 (556)	.07 (238)	.15 (129)	.14 (114)	.13 (74)	.45 (1)
Black spruce	− .01 (233)	.16 (700)	.17 (556)	.19 (27)	.26 (40)	.02 (75)	.29 (2)
Tamarack	− .51 (4)	.03 (227)	.08 (186)	− .24 (21)	− .33 (15)	− .06 (4)	.89 (1)
N. white-cedar	− .02 (236)	.37 (884)	.26 (321)	.47 (104)	.44 (184)	.37 (225)	.55 (50)
Hemlock	.04 (52)	− .13 (296)	—	.06 (46)	− .29 (118)	− .05 (117)	.01 (15)
Black ash	.45 (26)	.16 (163)	.06 (41)	.38 (48)	.03 (29)	.27 (7)	.07 (38)
Red maple	− .01 (152)	− .05 (742)	− .18 (61)	.06 (177)	− .05 (59)	− .13 (243)	.00 (202)
Elm	− .15 (103)	.24 (170)	—	.31 (44)	.13 (110)	.47 (8)	1.17 (8)
Yellow birch	.14 (65)	.00 (232)	.76 (2)	− .02 (81)	.06 (72)	− .07 (65)	− .04 (12)
Basswood	− .07 (231)	.38 (243)	—	.28 (77)	.36 (135)	.08 (5)	.88 (26)
Sugar maple	− .03 (866)	.08 (777)	—	.19 (266)	− .03 (305)	.00 (151)	.45 (55)
White ash	− .08 (193)	.35 (60)	—	.20 (18)	.02 (12)	—	.58 (30)
White oak	− .16 (63)	.08 (244)	—	—	—	—	.08 (244)
Select red oak	− .11 (136)	.15 (246)	− .15 (2)	.20 (111)	− .08 (24)	− .07 (24)	.22 (85)
Other red oak	− .07 (4)	.76 (154)	—	—	—	—	.76 (154)
Hickory	− .23 (54)	—	—	—	—	—	—
Bigtooth aspen	− .17 (5)	− .04 (184)	.32 (25)	− .22 (54)	− .23 (26)	− .18 (30)	.16 (49)
Quaking aspen	.07 (285)	.27 (662)	.33 (252)	.07 (106)	.21 (171)	.37 (93)	.44 (40)
Paper birch	.01 (130)	.10 (1,068)	.16 (726)	− .05 (97)	− .19 (122)	.17 (94)	.18 (29)
All[1]	− .03 (7,702)	.11 (11,182)	.06 (4,572)	.13 (1,587)	.09 (1,775)	.09 (1,720)	.28 (1,528)

[1]Includes all species on the property.

Table 3.—*Summary of the mean number of trees errors (trees/acre) by Lake States forest types in 10 years using the STEMS model. The number of plots appears in parenthesis. Positive values are overpredictions.*

Forest type	Calibration		Validation													
			All		Cloquet		Chequamegon		Nicolet		Hiawatha		Manistee	
Jack pine	45	(40)	7	(127)	4	(79)	14	(7)	72	(6)	15	(20)	-19	(15)
Red pine	3	(89)	6	(85)	1	(49)	60	(6)	-24	(4)	-1	(11)	12	(15)
White pine	-1	(12)	1	(13)	16	(6)	0	(2)	-44	(2)	1	(3)	—	
White spruce	-56	(2)	2	(10)	5	(4)	0	(1)	2	(4)	-3	(1)	—	
Balsam fir	7	(3)	31	(51)	16	(23)	38	(14)	-4	(8)	121	(6)	—	
Black spruce	15	(6)	15	(40)	19	(29)	-182	(2)	83	(3)	26	(6)	—	
Tamarack	—		-17	(23)	-8	(19)	-2	(2)	-116	(2)	—		—	
N. white-cedar	-32	(4)	-23	(47)	15	(14)	-18	(8)	-55	(9)	-59	(14)	89	(2)
Hemlock	-3	(2)	1	(7)	—		—		9	(4)	-9	(3)	—	
Lowland hdwds.	15	(6)	32	(30)	21	(3)	75	(5)	15	(2)	-139	(1)	33	(19)
N. hardwoods	-3	(76)	-2	(135)	—		14	(40)	-20	(49)	1	(28)	9	(18)
White oak	-3	(3)	-3	(15)	—		—		—		—		-3	(15)
Northern red oak	2	(10)	68	(15)	—		116	(4)	95	(1)	58	(2)	44	(8)
Oak-pine	—		-33	(10)	—		—		—		-16	(1)	-35	(9)
Oak-hickory	1	(6)	-36	(35)	—		-300	(1)	—		—		-28	(34)
Aspen	2	(31)	-4	(121)	5	(30)	26	(23)	-38	(42)	-6	(14)	31	(12)
Paper birch	5	(3)	5	(58)	3	(36)	-21	(8)	8	(9)	76	(4)	2	(1)
All	6	(293)	2	(822)	6	(292)	16	(123)	-18	(145)	5	(114)	2	(148)

four properties than they were for either the calibration data or Cloquet. If the mortality function is accurate, these deviations should average out over all properties. This can be evaluated by checking the "all validation" results.

Five forest types are in error by more than 30 trees in 10 years—balsam fir (31), lowland hardwoods (32), oak-pine (-33), oak-hickory (-36), and northern red oak (68). Limited data were originally available to calibrate these particular forest types, and note that three of them include the oak species. When looking at the individual property results, northern red oak is clearly the worst. Even though jack pine had a large error (45) on the calibration data, it did well on the validation data.

Overall model (BA errors). The BA errors provide information on the complete model, i.e., how well the growth and mortality components work together (table 4). For all the validation properties, the forest types with BA errors more than 9 square feet in 10 years are balsam fir (9.5), black spruce (9.4), northern white-cedar (9.1), and lowland hardwoods (19.6). The marked overestimation of lowland hardwood species is also evident on the calibration data. In all

Table 4.—*Summary of the mean basal area errors (sq. ft./acre) by Lake States forest types in 10 years using the STEMS model. The number of plots appears in parenthesis. Positive values are overpredictions.*

Forest type	Calibration		Validation													
			All		Cloquet		Chequamegon		Nicolet		Hiawatha		Manistee	
Jack pine	6.9	(40)	2.7	(127)	2.0	(79)	4.4	(7)	5.2	(6)	4.0	(20)	2.9	(15)
Red pine	-.9	(89)	-.9	(85)	-.2	(49)	5.0	(6)	-5.6	(4)	-8.0	(11)	.8	(15)
White pine	-3.3	(12)	.4	(13)	3.5	(6)	-.7	(2)	-7.1	(2)	-.0	(3)	—	
White spruce	-9.5	(2)	.5	(10)	4.3	(4)	-2.3	(1)	-.4	(4)	-7.9	(1)	—	
Balsam fir	7.7	(3)	9.5	(51)	10.8	(23)	4.2	(14)	5.5	(8)	22.1	(6)	—	
Black spruce	6.7	(6)	9.4	(40)	9.3	(29)	-1.8	(2)	25.4	(3)	5.9	(6)	—	
Tamarack	—		.0	(23)	1.4	(19)	1.0	(2)	-14.0	(2)	—		—	
N. white-cedar	-8.9	(4)	9.1	(47)	11.9	(14)	5.1	(8)	5.3	(9)	7.3	(14)	35.4	(2)
Hemlock	-6.9	(2)	-8.2	(7)	—		—		-8.5	(4)	-7.7	(3)	—	
Lowland hdwds.	10.3	(6)	19.6	(30)	10.2	(3)	22.0	(5)	6.0	(2)	-10.6	(1)	23.5	(19)
N. hardwoods	-.4	(76)	3.4	(135)	—		4.9	(40)	.7	(49)	-1.5	(28)	15.3	(18)
White oak	-3.7	(3)	-.9	(15)	—		—		—		—		-.9	(15)
Northern red oak	5.2	(10)	7.4	(15)	—		10.1	(4)	25.0	(1)	1.8	(2)	5.3	(8)
Oak-pine			-2.5	(10)	—		—		—		-22.1	(1)	-.3	(9)
Oak-hickory	3.4	(6)	-8.1	(35)	—		-15.0	(1)	—		—		-7.8	(34)
Aspen	-2.6	(31)	3.1	(121)	6.5	(30)	4.0	(23)	-1.2	(42)	1.2	(14)	10.2	(12)
Paper birch	5.9	(3)	3.3	(58)	3.3	(36)	-.3	(8)	-.7	(9)	17.9	(4)	9.2	(1)
All	.6	(293)	3.5	(822)	4.2	(292)	4.7	(123)	.8	(145)	2.2	(114)	5.0	(148)

problem cases mentioned above limited data were available to calibrate the model.

Error patterns. Error patterns in the growth component for the three key variables—DBH, NT, and BA—over various initial conditions shows that the problem areas in the validation data generally occur where there are few observations (figs. 3 and 4). Errors in these classes should cause little difficulty unless projections are to be made in these extreme ranges. Because much more variability exists in classes with few observations, caution needs to be exercised when emphasizing these trends.

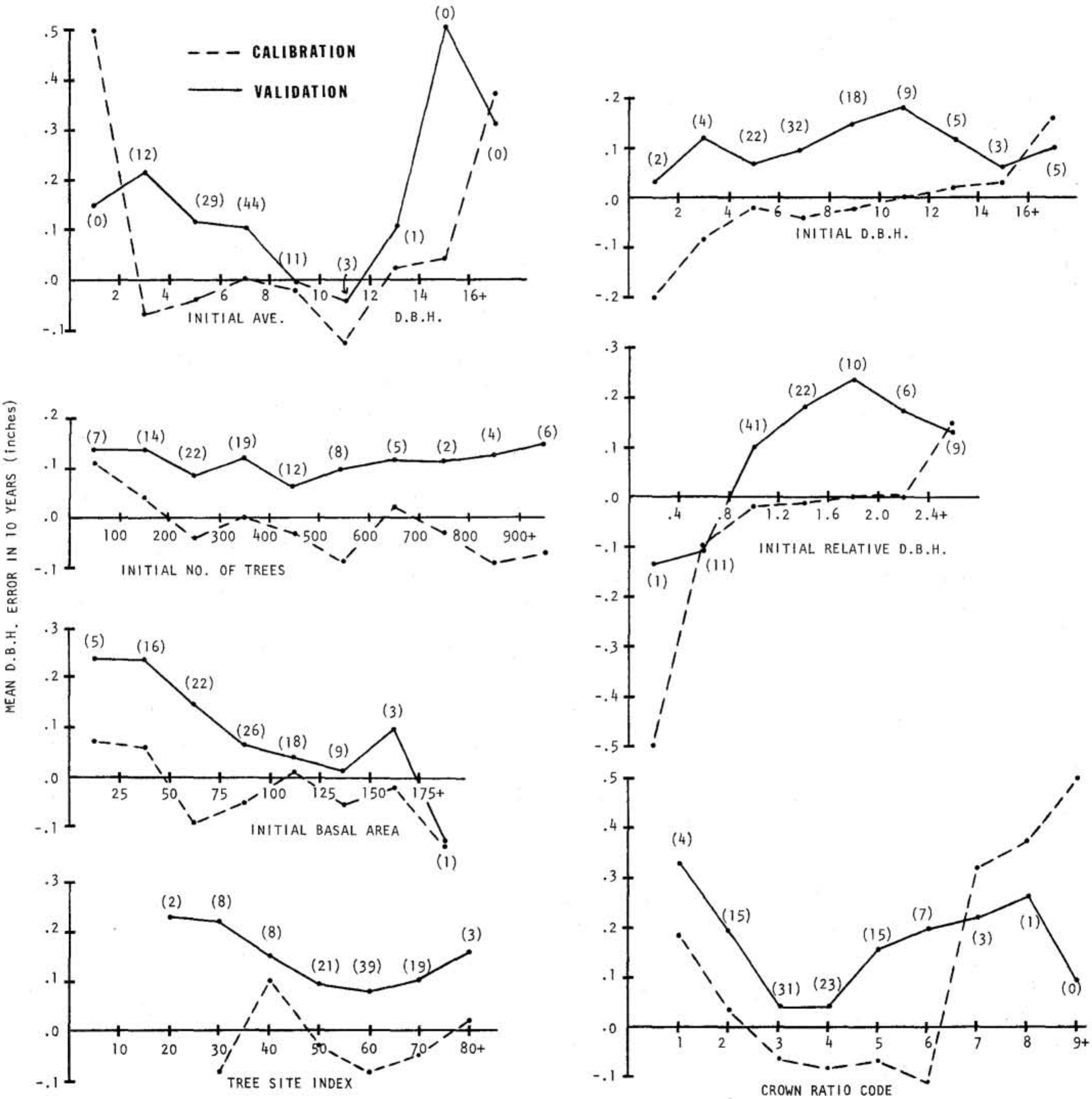


Figure 3.—*The mean d.b.h. errors in 10 years plotted against selected initial tree and stand conditions. The percent of the observations in each class for the validation data is given in parentheses.*

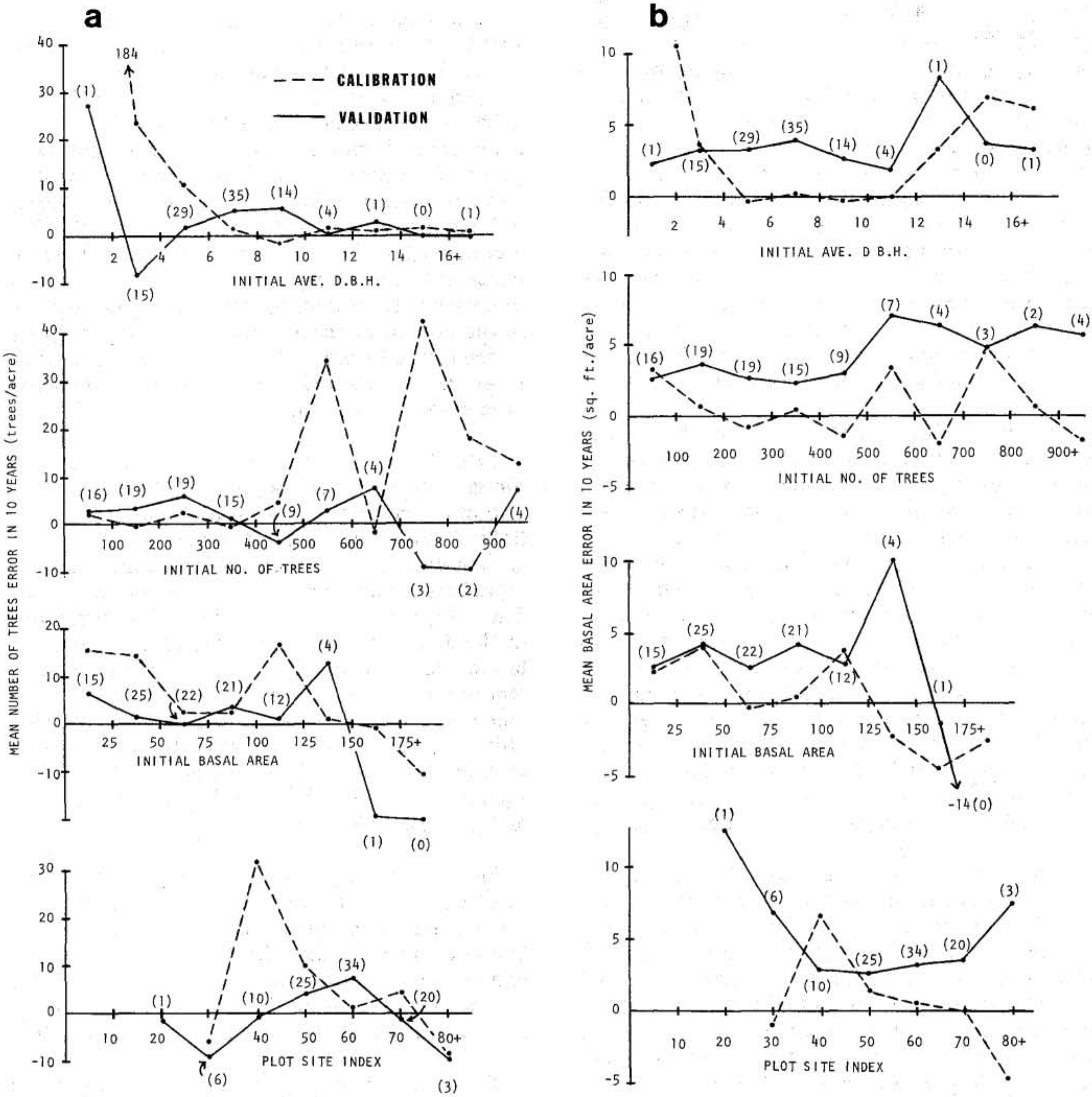


Figure 4.—*The (a) mean number of trees errors and (b) the mean basal area errors in 10 years plotted against selected initial stand conditions. The percent of the plots in each class for the validation data is given in parenthesis.*

It would seem logical to include age in the list of initial conditions evaluated. However, age is not always an accurate variable. For example, when a plot has been clearcut, the new age becomes zero even though a few undesirable old trees may still remain. Thus a plot listed as 5 years old could have an average diameter of 12 inches. Consequently we have not included age. Many of the same results can be obtained by studying average diameter trends.

Some potential problem areas surfaced on both the calibration and validation data. One such area is the excessive growth put on trees in stands with large initial average diameters and to a lesser degree on trees in stands with very low average diameters (fig. 3). This evidence is based on few trees. On the initial basal area graph, there is a very slight trend to overgrow low basal area stands on the calibration data. This trend is clearly magnified on the validation data. (Recall that we had much more data in the low basal area range on the validation properties (fig. 2).) For tree site index a noticeable curvilinear trend appears in the errors for the validation data, although, because of the distribution of the data, the model performs well for 79 percent of the trees. This same trend also appears in the calibration data, but at a mean DBH error of 0.2 inches lower. Likewise, a curvilinear trend similar to that for tree site index appears for crown ratio code. Tree site index and crown ratio code are handled in the same term of the growth potential function so these results indicate that the functional form chosen does not adequately describe the biological mechanisms at work.

The initial relative d.b.h. (relative d.b.h. = d.b.h./average d.b.h.) for the validation properties shows a clear increasing trend with a marked overgrowth of overstory trees (relative d.b.h. greater than 1). This trend is bothersome because even a small overgrowth will be magnified with longer projections. The trend tapers off for very large relative values and so becomes somewhat self-correcting. Because this trend is not evident for the same class of trees on the calibration data, the problem is likely due to a difference between the data bases. "Extreme" understory trees (i.e., relative d.b.h. 0.4 or less) are sharply underestimated on the calibration data and somewhat underestimated on the validation data. This would account for the slight underestimation of DBH for trees with small initial diameters.

The problem areas are due to difficulty with the form of the equations used to estimate the various growth components or to a limited amount of data to calibrate against. The main problems indicate possible model weaknesses in how the crown ratio and site index are incorporated into the growth potential and how competition (as measured by basal area and possibly relative diameter) is handled in the modifier function. However, the various parts of the growth component are so interrelated that correcting one problem often introduces difficulties with another part of the model. Those cases where difficulties are clearly evident occur from 10 to 20 percent of the time, often under somewhat extreme conditions.

No trends emerge from the error patterns for the mortality component when evaluating both the validation and calibration data (fig. 4a). However, NT variability increases for stands with low average d.b.h. and with high number of trees. Both results are logical in that a greater number of trees will magnify NT errors.

Finally, performance for the entire model, as measured by the mean BA error (fig. 4b), shows no clear trends. However, errors in the growth and mortality components may cancel each other or combine to heighten an already noticeable trend. We investigated several examples of error effects and found that a 10 percent error in DBH produced roughly a 20 percent error in BA, whereas a 10 percent error in NT produced only a 10 percent error in BA. Thus accuracy in predicting diameter growth is twice as important as accuracy in predicting the mortality component. In evaluating the data, trends evident in the growth component have been largely muted by the mortality effect. The one exception is the site index trend for the validation properties, which seems even more pronounced.

DBH error by diameter class. Further analysis of the growth component is provided through evaluation of diameter errors for initial diameter classes (table 5). For many species the model estimates tree growth well for the lower diameter ranges but overpredicts growth for the middle diameter classes. This overprediction is smaller for the few larger trees in the data base. When the results are graphed, a few species show a marked relation between diameter and diameter prediction error. We found clear increasing trends for white pine, northern white-cedar, and black ash and strong decreasing trends for bigtooth aspen and quaking aspen. The red pine equation consistently underpredicted growth on trees less than 14 inches d.b.h.

Species mortality rates. The mortality component can be evaluated as a tree attribute in addition to the NT evaluation as a stand attribute. As a tree attribute, we calculated the observed and predicted annual mortality rates for each species based on the number of trees that survive through an observation interval and the length of that interval (table 6).

The observed mortality rates vary widely from species to species. For example, the observed annual mortality rate for quaking aspen is more than 30 times as great as that for red pine. More important, however, is how close the predicted rate is to the observed rate. In many cases it is very close—red pine, hemlock, red maple, and yellow birch to name a few. Several, however, are not close—balsam fir, black spruce, elm, tamarack, black ash, and other red oaks. Because the last three of these were calibrated from limited data, the function coefficients may not be reliable. The remaining three—balsam fir, black spruce, and elm—are subjects of insect or disease outbreaks. Balsam fir is defoliated by the

Table 5.—*Mean d.b.h. errors (inches) in 10 years by species and diameter class for all validation data sources. Positive values are overpredictions.*

Species	No. of trees[1]	Diameter class (inches)								All classes	Mean d.b.h. growth in 10 years
		0-3+	4-5+	6-7+	8-9+	10-11+	12-13+	14-15+	16+		
Jack pine	1,539	−0.12	0.05	0.08	0.21	0.27	0.21	0.14	−0.65	0.13	1.01
Red pine	1,360	−.33	−.12	−.20	−.44	−.43	−.14	.01	.45	−.17	1.72
White pine	297	−1.00	−.42	−.38	−.11	.20	.49	.12	.17	−.01	1.78
White spruce	148	.37	.08	.27	.43	.61	.46	.49	.55	.32	1.65
Balsam fir	556	−.05	.02	.17	.33	.31	.34	.41	—	.11	1.13
Black spruce	700	.03	.15	.18	.22	.32	—	−2.80	—	.16	.73
Tamarack	227	−.36	.02	−.02	.05	.47	−.06	.53	—	.03	.67
N. white-cedar	884	.09	.19	.31	.51	.65	.82	.64	.52	.37	.82
Hemlock	296	−.17	−.22	−.04	−.12	.06	−.29	−.14	−.15	−.13	.94
Black ash	163	−.35	−.02	.15	.35	.54	.51	.30	.27	.16	.61
Red maple	742	.29	−.01	−.07	−.06	−.15	−.20	−.51	−.29	−.05	1.20
Elm	170	−.19	−.01	.20	.43	.28	.05	.71	.36	.24	1.39
Yellow birch	232	.04	.09	.02	−.21	.20	.10	−.03	−.04	−.00	.92
Basswood	243	.50	.70	.38	.29	.26	.20	.38	.41	.38	1.31
Sugar maple	777	.24	.09	.06	.10	.21	.06	.12	−.23	.08	1.20
White oak	244	−.12	.10	.10	.19	.11	.09	.02	−.05	.08	.89
Select red oak	246	.25	.30	.17	.05	.14	−.19	−.08	.31	.15	1.46
Other red oak	154	—	.96	.54	.95	.88	.83	.84	.50	.76	1.12
Bigtooth aspen	184	.34	.27	.11	−.15	−.37	−.71	−.86	−.96	−.04	1.55
Quaking aspen	662	−.13	.24	.38	.33	.16	−.10	−.20	−.73	.27	1.45
Paper birch	1,068	−.14	.02	.11	.21	.15	.34	.25	.03	.10	1.07
All[2]	11,182	.09	.07	.10	.15	.18	.12	.06	.10	.11	1.16
Percent trees in diameter class		6	22	32	18	9	5	3	5	100	

[1]Species with fewer than 100 observations have been omitted.
[2]Includes species with fewer than 100 observations.

Table 6.—*Predicted and observed annual tree mortality rates by Lake States species for all validation properties (includes all trees that were initially alive and not cut).*

Species	No. of trees	Annual tree mortality rate	
		Predicted	Observed
		Percent	
Jack pine	2,336	1.2	1.6
Red pine	1,397	.1	.1
White pine	363	.3	1.1
White spruce	187	1.2	.6
Balsam fir	1,092	1.4	3.2
Black spruce	1,024	.6	1.8
Tamarack	370	2.5	.6
N. white-cedar	987	.3	.6
Hemlock	323	.3	.5
Black ash	254	.9	2.4
Red maple	852	.7	.6
Elm	293	1.2	3.8
Yellow birch	321	1.6	1.6
Basswood	303	1.2	1.0
Sugar maple	946	1.4	.7
White ash	65	.1	.6
White oak	274	.4	.7
Select red oak	273	.6	.5
Other red oak	312	5.8	.8
Bigtooth aspen	270	1.7	1.5
Quaking aspen	1,468	2.5	3.3
Paper birch	1,191	.2	.5
All[1]	15,240	1.1	1.4

[1]Includes all species on the property.

spruce budworm, *Choristoneura fumiferana;* black spruce is a host for dwarf mistletoe, *Arceuthobium pusillum;* and elm is killed by Dutch elm disease caused by the fungus *Ceratocystis ulmi* (USDA Forest Service 1979b). None of these damaging agents were present at high levels on the calibration locations during the measurement periods involved. In several of the validation areas, however, these agents were present at high levels.

Balsam fir mortality is underpredicted on all the validation properties. For all four properties where balsam fir is an important component, spruce budworm was present in the area during the measurement period (USDA Forest Service 1962, 1968, 1971, 1977). Therefore, spruce budworm may partially account for the poor performance of the balsam fir mortality predictions.

Most of the black spruce data is from the Cloquet forest. However, dwarf mistletoe has not caused significant mortality on the Cloquet forest (D. W. French and F. A. Baker, personal communications). So mistletoe does not account for the discrepancy.

Predicted and observed elm mortality rates are close for all sources except the Manistee National Forest. There, the observed mortality rate was 18 percent as compared to a predicted rate of only 1 percent. Dutch elm disease was responsible for large elm losses in lower Michigan in 1962 and by 1971 most of the elms were dead (USDA Forest Service 1962, 1972). Yet it wasn't until near the end of the measurement period that the disease was common in northern Wisconsin and upper Michigan (USDA Forest Service 1977). This helps to explain the contrast between Manistee and the other areas.

Precision analysis

In this section we will discuss some general aspects of variablity in the model evaluation process. The first analysis is a comparison of the actual and predicted distribution for the three test variables: DBH, NT, and BA (fig. 5). All three test variables agree closely between the predicted and observed distributions with perhaps a bit more discrepancy in the basal area results, which is to be expected. The closeness of these two distributions is in part a function of the interval length. The longer the projections, the greater the discrepancies no matter how well the model predicts. These results are standardized to 10 years for all validation properties combined.

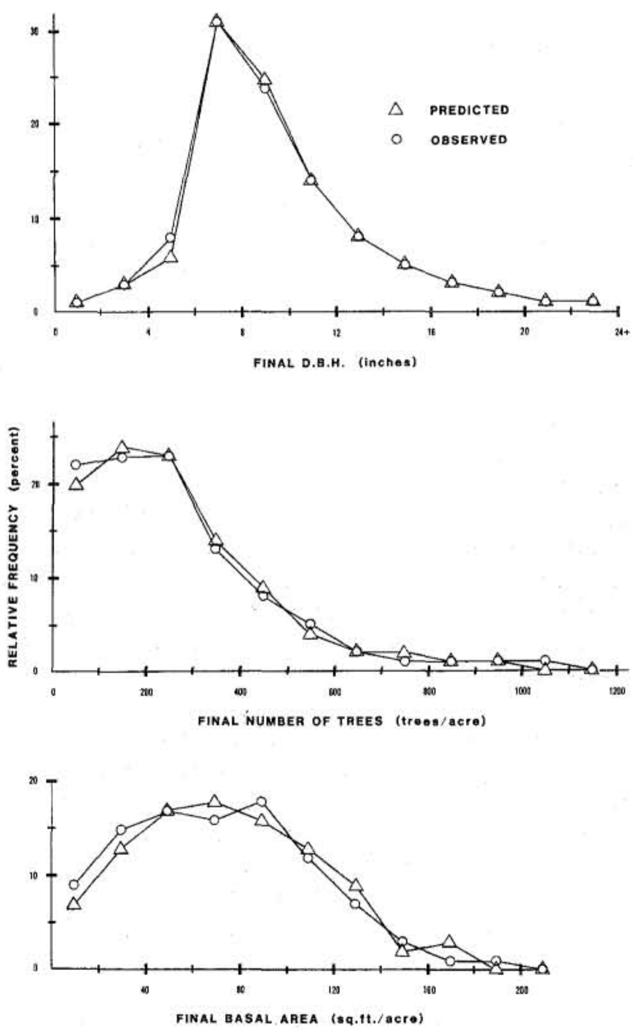


Figure 5.—*A comparison of the frequency distributions of the predicted and observed values for tree d.b.h., number of trees, and stand basal area, standardized to 10 years. All validation sources are included.*

The second analysis of variability is the cumulative frequency of the errors for each test variable (fig. 6). Differences are shown between predicted and observed values for DBH, NT, and BA standardized to 10 years for all validation properties.

The graphed results can be viewed two ways. One is to say that 68 percent of the tree diameter observations will have an error of between -0.35 and 0.55 inches in 10 years (or, 95 percent are between -1.2 and 1.2 inches). The other is to know what percent of the trees had predicted DBH values within, for example 0.5 inch of the observed DBH. Viewed in this manner, 71 percent of the trees (82 percent–11 percent) had predicted diameters within 0.5 inches of the true value, 88 percent of the plots (93 percent–5 percent) had predicted NT within 100 trees per acre, and 68 percent of the plots (78 percent–10 percent) had BA values predicted to within 10 square feet.

A third precision analysis, based on the Cloquet Experimental Forests, is included in the time analysis discussion below.

Time analysis

Short-term effect. Prediction errors can be expected to increase as the length of the projection interval increases. We examined these error trends to determine their magnitude and form.

We used the data from the Cloquet validation for the time analysis because it was the only one that was remeasured more than once after the initial measurement. It was remeasured 5, 10, and 17 years after the initial measurement. The analysis was done at the stand level using BA errors, which provide a good overview of the entire model performance on a variable of interest to most users. On the average, we found that errors increase as the projection interval increases and variability of the errors increases as well (fig. 7).

An analysis of accuracy for all types combined shows that BA is overestimated on the average by

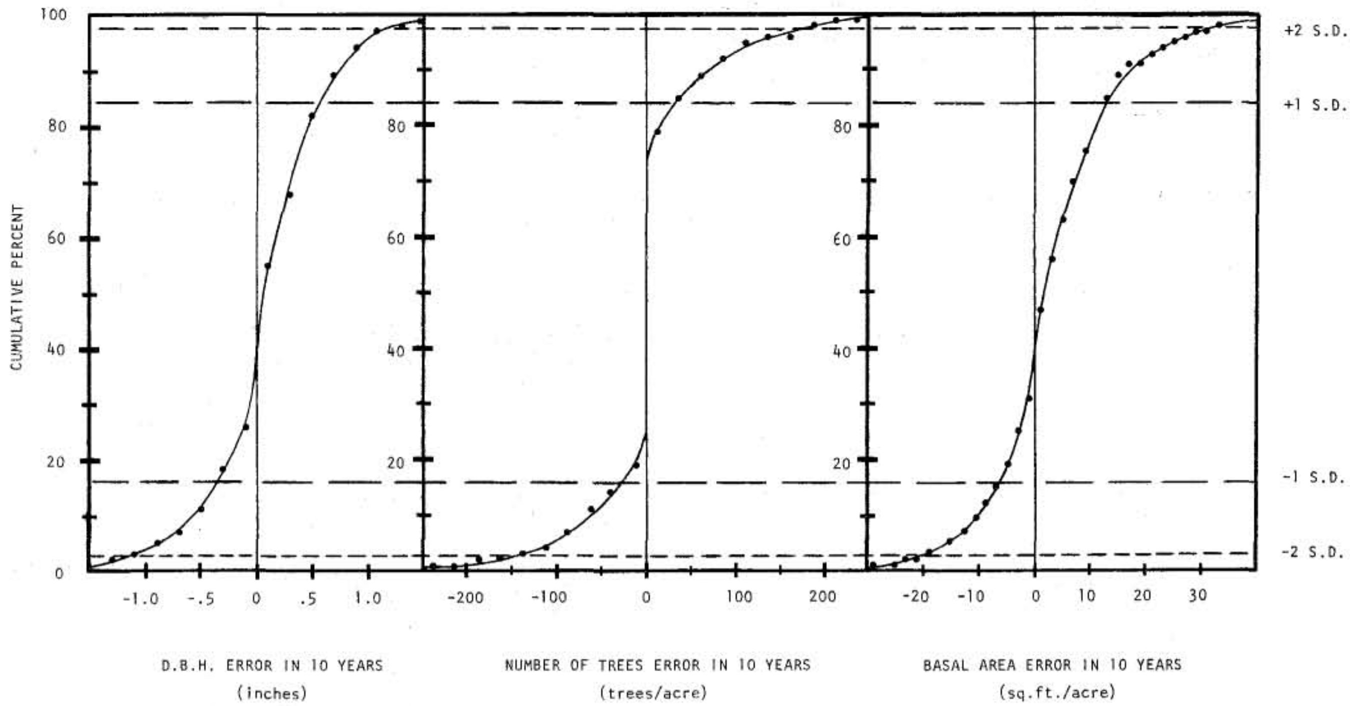


Figure 6.—*Cumulative frequency of 10 year errors in predicting tree d.b.h., number of trees, and stand basal area for all validation sources. The error equals the predicted attribute minus the observed attribute. Lines representing ±1 and ±2 standard deviations have been added.*

7 square feet in 17 years. However, some types (balsam fir, black spruce, northern white-cedar, and lowland hardwoods) were overestimated by as much as 16 to 20 square feet. These types account for 24 percent of all the plots.

The precision results for all types show the standard deviation of the BA error to be 16 square feet in 17 years. Stated in slightly different terms, 68 percent of the basal area errors fall between -6 and 13 square feet after 10 years and between -9 and 23 square feet after 17 years. When the mean error and/or standard deviation are large, fairly high BA errors may be encountered (e.g., balsam fir, black spruce, and quaking aspen types). Atypical patterns were found for white spruce and lowland hardwoods.

But we still need to assess the assumption of linear decay of errors made when the time adjustment was introduced. If the annual rate of decay is constant, the lines should be linear through the origin. When *all* types are combined, the linear assumption holds up to 10 years. Adjusting the 17 year values to 10 years would usually result in a slight overestimation of BA error at 10 years because the pattern is slightly convex. Further analysis, not given here, revealed that NT trends were more convex. Therefore, on the average, the stand level results when adjusted to 10 years will make STEMS appear worse than it is. However, at the tree level combining all species, the DBH trends were nearly linear.

A few of the individual forest types on Cloquet had noticeable nonlinear trends (fig. 7). But the only clear cases in which the error would be *underestimated* are white pine (6 plots), white spruce (4), and balsam fir (23). The remaining four validation properties had measurement lengths of 9 to 11 years. Adjusting these results to 10 years presents no problem. This supports our adjustment approach, especially on the validation properties.

Long-term effect. An evaluation of long-term projections, as used when studying forest succession, becomes more difficult because there are no existing remeasured plots for comparison. One evaluation approach is to compare the results of the projection with expectations based on professional judgment. This is often sufficient to detect gross errors in the model. Another approach is to compare the results with published yield tables. However, there are problems with this approach: (1) the yields are based on summaries of different stands at specific ages; (2) the averages derived from these stands represent the development of a *theoretical* stand; (3) the results do not represent an actual growth series; and (4) it is difficult to de-

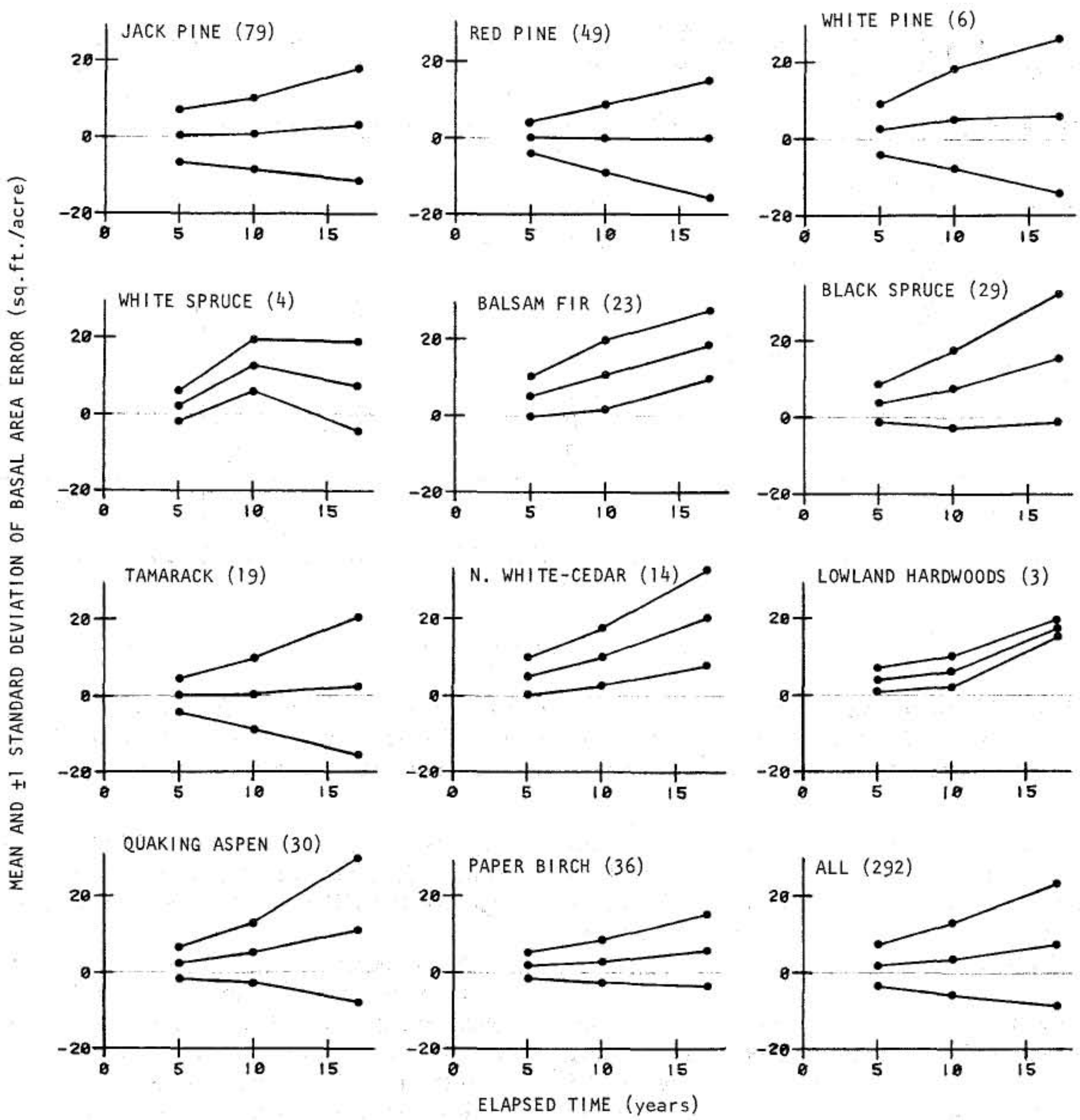


Figure 7.—*Pattern of mean and standard deviation of basal area error by forest type on the Cloquet Experimental Forest, Minnesota. Standard deviations are shown by the outer lines. The number of plots for each forest type is given in parentheses.*

velop an appropriate initial tree list because the yield tables do not specify such things as species composition and diameter distributions. Despite these problems, we used a combination of scientific judgment and published yield tables to evaluate the model's long-term predictive capabilities.

We made two long-term projections on pure stands—a conifer (red pine) and a hardwood (quaking aspen) (fig. 8). Both stands, site index 60, were begun at age 20 and were projected for 100 years. The red pine stand projections showed good agreement with the basal area table of Stiell and Berry (1973) at age 50, the extent of the published table. For quaking aspen, the greatest difference between the projection and the yield table values of Brown and Gevorkiantz (1934) was 18 square feet per acre at age 50. Both projections followed the general trends indicated from the yield tables and continued on in a reasonable manner. Although this is only a cursory examination of the model's long-term projections, it suggests that the model does not produce gross errors for red pine and aspen.

Data base differences

The calibration and validation data bases have one very important difference. The calibration data came from research plots so the coefficients were

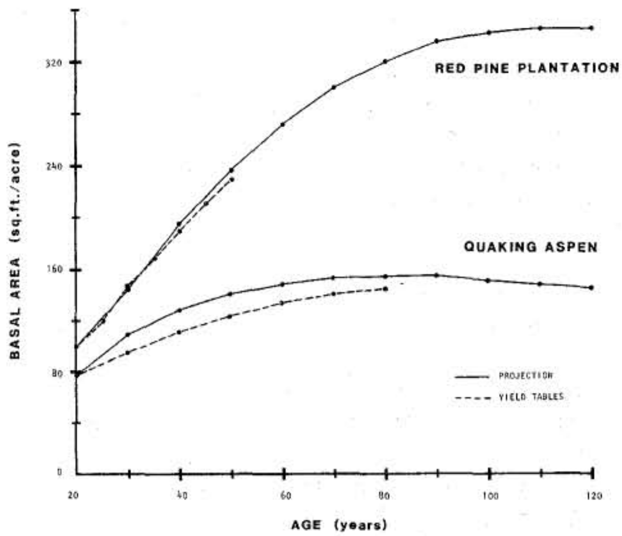


Figure 8.—*Comparison of long-term STEMS projections with published yield table results for red pine and quaking aspen stands on a site index of 60.*

developed primarily from plots where uniform spacing was the goal. However, the validation data came from general plots so the model was applied on plots where spacing may be far from ideal. Unequal spacing may reduce production below the optimum found with uniform spacing. The expected results, when projecting general stands, would be to overpredict growth. Certainly, diameter growth is consistently overestimated on all validation properties (table 2). These results agree with those of Bruce (1977).

Also the range of conditions differs between the calibration and validation data bases. The data base used to calibrate the model contained fewer low basal area plots and more high basal area plots than the validation properties (fig. 2). Our results (fig. 3) show that the model overgrows low basal area stands. Therefore, if this type of stand is predominant, the model would perform poorly. This is also probably a contributing factor to the overprediction on the validation properties.

Crown ratio analysis

Crown ratio, an important variable in the growth potential, had been measured on each tree used in the evaluation. We analyzed the growth and mortality components of the models without introducing errors from crown ratio estimates. Therefore, we will present a separate validation of the crown ratio function for those cases in which crown ratio would be estimated in STEMS, i.e. not recorded when trees were measured. Because all trees in the validation data had a known crown ratio, we used the crown ratio function to predict the crown ratio of each tree and then we evaluated the resulting prediction errors.

The overprediction of crown ratio for all species combined is 0.78, with individual species ranging from 0.31 to 1.30, except for red pine at -0.07 (table 7). Plotting the mean crown ratio code errors over stand basal area also provides evidence of a marked basal area trend in which the errors decrease as basal area increases (fig. 9).

The crown ratio function in STEMS was calibrated using Forest Survey inventory data from Wisconsin and northern Minnesota[6] that included a wide range of stand basal area and diameter conditions. The crown ratio codes observed from the inventory data tended to be as much as 1 unit higher than those observed on the calibration data. We found no apparent cause for this discrepancy, but obviously it has contributed to the overestimates of crown ratio.

How much effect do these overprediction errors have on the estimated growth? One National Forest, Chequamegon, was also projected assuming unknown crown ratios. The results showed that when the crown ratio function is used to estimate the crown ratios, the mean DBH error increased from 0.13 to 0.25 inches in 10 years and the mean BA error increased from 4.7 to 8.1 square feet per acre in 10 years.

[6]*See page 7.*

Table 7.—*Summary of crown ratio function test showing mean crown ratio code errors by Lake States species for all trees on the validation properties.*

Species	No. of trees	Mean crown ratio code error
Jack pine	1,906	1.07
Red pine	1,626	−.07
White pine	368	1.07
White spruce	257	.79
Balsam fir	988	1.10
Black spruce	957	1.26
Tamarack	385	1.02
N. white-cedar	1,002	.44
Hemlock	306	.31
Red maple	837	1.09
Elm	194	1.01
Yellow birch	272	1.02
Basswood	274	1.30
Sugar maple	878	1.18
White ash	61	1.30
White oak	255	.52
Select red oak	259	.57
Other red oak	286	.44
Bigtooth aspen	235	.94
Quaking aspen	1,140	.94
Paper birch	1,277	.31
Total	13,763	.78

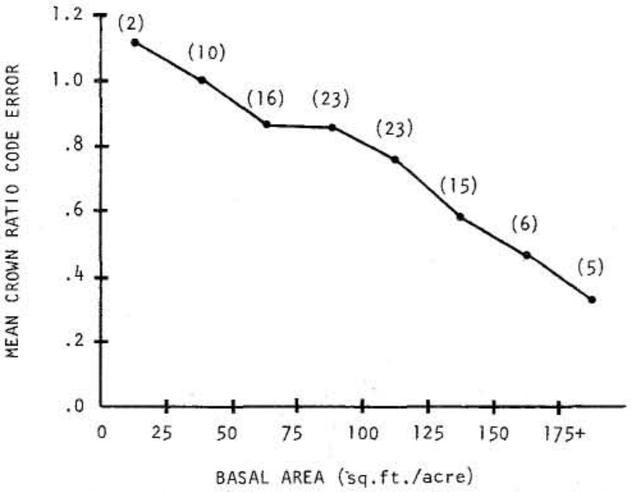


Figure 9.—*Mean crown ratio code errors plotted against the stand basal area for all trees on the validation properties. The percent of the observations in each class is given in parentheses.*

It is possible that crown ratios vary more by location than other variables. Therefore, users may have to test the predicted crown ratio values against known values from the location of interest to determine if an adjustment is needed. At present, if the STEMS model is used to grow trees with unknown crown ratios, it will probably overpredict crown ratios and hence overgrow tree diameters.

SUMMARY

A projection system is evaluated by analyzing its capabilities to predict behavior in the real world. In this paper, we evaluated the STEMS projection system by analyzing the differences (or errors) between the predicted and observed values on three key variables—tree d.b.h. (DBH), number of trees per acre (NT), and stand basal area per acre (BA). Performance of the growth component was measured by DBH errors, the mortality component by NT errors, and the system as a whole by BA errors. Many of the basic tests were performed on two data sets: (1) a systematic sample of every fifth plot from the calibration data base, and (2) an independent data base comprised of five different sources located in Minnesota, Wisconsin, and Michigan.

Initially, we evaluated two growth models—the original GTR49 model and the current STEMS model. The results indicate the two models are similar in many ways. However, the STEMS model (1) is slightly more precise; (2) performs slightly better than the GTR49 model in three areas as judged by error patterns; (3) does not grossly overpredict yields on long-term projections as the GTR49 model did on some species; and (4) is simpler to calibrate, understand, and use.

Included in the tests on STEMS are analyses of accuracy, precision, and time effect. The important findings and observations are summarized below, including an attempt to identify the major areas of weakness of the STEMS model.

1. Both models predicted growth and mortality well on the calibration data base. Therefore, we have no reason to suspect any basic mathematical flaw in either model.
2. STEMS did reasonably well projecting plots from the five independent properties comprising the validation data source. The *major* problem areas involved predicting diameter growth for other red oak, mortality for northern red oak, and stand basal area for lowland hardwoods. In most of these cases, few trees were available to use for calibration.
3. In an evaluation of the error patterns for the three test variables over a wide range of initial conditions, we found that the STEMS model overgrows trees under the following conditions:
   (a) when basal area of the stands are low—especially below 50 square feet per acre,
   (b) when tree site index is 30 and below,
   (c) when crown ratio codes are 2 and below, and
   (d) when the relative d.b.h. of overstory trees is between 1.2 and 2.2.

   The problems listed in (b) and (c) indicate possible model weaknesses in the growth potential and those in (a) and (d) with competition as it is handled in the modifier function.

4. Tree diameter errors given by species and initial diameter classes generally show that the model predicts diameter growth well for the lower diameter classes but overpredicts growth for the middle diameter classes. This overprediction decreases for the few larger trees in the data base.
5. The problem with underpredicting mortality on balsam fir and elm on the validation properties may be due to outbreaks of spruce budworm and to Dutch elm disease. These damaging agents were not present at high levels on the calibration locations during the measurement periods involved.
6. In general, very little error was introduced in the validation results by standardizing the time and if errors do result, they usually will be conservative. This approach to model testing allows us to use and compare data with different measurement intervals. But standardization of the time is only valid if the decay of errors is approximately linear for the measurement length used.

In the Cloquet example given, most species showed fairly linear trends through 17 years. In those cases where the pattern was slightly convex, adjusting to 10 years results in a slight overestimation of the error. The other four validation properties had measurement intervals of 9 to 11 years, so we had no problem with those data.

7. The STEMS model was tested using long-term projections for two species. The projections seemed reasonable when compared with results from yield tables.
8. Diameter growth was consistently overpredicted on all validation properties. We are not so much concerned with the magnitude of the error as with the possibility of a consistent positive bias. Is there a possible explanation for these results? One possible explanation is that research plots were used to calibrate the model. Growth on these plots would be greater because they are uniformly spaced and mortality would be lower because of the cutting procedures used. Another possible explanation is that the validation properties had a large proportion of low basal area plots. These plots were consistently overgrown, which probably contributed to the overprediction.
9. Finally, when the crown ratio function was used to estimate missing crown ratio codes, it overpredicted them by an average of 0.78 units. This will ultimately overpredict the diameter growth as well. If crown ratio is missing, an adjustment factor may be necessary to correct for this bias in the function.

In conclusion, on the validation properties the STEMS model overgrows trees on the average by 0.11 inches in 10 years. On 71 percent of these trees it predicted the diameter to within 0.5 inches of the observed diameter after 10 years growth. Likewise, the model overestimated stand basal area on the average by 3.5 square feet per acre after 10 years growth and on 70 percent of these plots it predicted the basal area to within 10 square feet per acre of the true value.

This paper provides the users of the STEMS projection system with extensive quantitative information about how well the model predicts and also highlights possible problem areas. Nevertheless, it is the user's responsibility to determine the model's usefulness for his particular application. His decision needs to be based on an in-depth examination of all the evidence. Careful scrutiny of the tables and figures is important. Furthermore, we encourage users to test the model on their own data whenever possible. This will help them determine if the model will meet their needs, or what parts of the model may be suspect, before making a major commitment of time and money.

LITERATURE CITED

Belcher, David M.; Holdaway, Margaret R.; Brand, Gary J. STEMS—the Stand and Tree Evaluation and Modeling System. Gen. Tech. Rep. NC-79. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1982. 18 p.

Brown, R. M.; Gevorkiantz, S. R. Volume, yield, and stand tables for tree species in the Lake States. Tech. Bull. 39. St. Paul, MN: University of Minnesota Agricultural Experiment Station; 1934. 208 p.

Bruce, D. Yield differences between research plots and managed forest. J. For. 75: 14-17; 1977.

Carmean, Willard H. Site index comparisons among northern hardwoods in northern Wisconsin and upper Michigan. Res. Pap. NC-169. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1979. 17 p.

Carmean, Willard H.; Vasilevsky, A. Site-index comparisons for tree species in northern Minnesota. Res. Pap. NC-65. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1971. 8 p.

Christensen, Linda; Hahn, Jerold T.; Leary, Rolfe A. Data Base. In: A generalized forest growth projection system applied to the Lake States region. Gen. Tech. Rep. NC-49. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1979. p. 47-55.

Goodall, David W. Building and testing ecosystem models. In: Jeffers, J. N. R., ed. Mathematical models in ecology. Oxford: Blackwell Sci. Publ.; 1972: 173-194.

Goulding, C. J. Validation of growth models used in forestry management. New Zealand J. For. 24: 108-124; 1979.

Leary, Rolfe A.; Hahn, Jerold T.; Buchman, Roland G. Tests. In: A generalized forest projection system applied to the Lake States region. Gen. Tech. Rep. NC-49. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1979. p. 79-89

Mankin, J. B.; O'Neill, R. V.; Shugart, H. H.; Rust, B. W. The importance of validation in ecosystem analysis. In: Innis, G. S., ed. New Directions in the Analysis of Ecological Systems. La Jolla, CA: Simulation Council, Part I: 63-71; 1975.

Pilgrim, D. H. Model evaluation, testing and parameter estimation in hydrology. In: Chapman, T. G.; Dunin, F. X., eds. Prediction in catchment hydrology, Proceedings, National Symposium on Hydrology; 1975 Nov. 25-27; 1975: 305-333.

Schaeffer, D. L. A model evaluation methodology applicable to environmental assessment models. Ecol. Modelling 8: 275-295; 1980.

Stiell, W. M.; Berry, A. B. Yield of unthinned red pine plantations at the Petawawa Forest Experiment Station. Publ. 1320. Department of the Environment, Canadian Forestry Service; 1973: 16 p.

Taylor, M. A. P. Evaluating the performance of a simulation model. Transportation. Res. 13A: 159-173; 1979.

USDA Forest Service. Forest pest conditions of the Central and Lake States. Milwaukee, WI: Division of state and private forestry; 1962. 7 p.

USDA Forest Service. Forest insect conditions in the United States, 1962. Washington, DC: 1968. 39 p.

USDA Forest Service. Forest insect conditions in the United States, 1970. Washington DC: 1971. 44 p.

USDA Forest Service. Forest insect and disease conditions in the United States, 1971. Washington DC: 1972. 65 p.

USDA Forest Service. Forest insect and disease conditions in the United States, 1974. Washington DC: 1977. 55 p.

USDA Forest Service. A generalized forest growth projection system applied to the Lake States region. Gen. Tech. Rep. NC-49. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1979a. 96 p.

USDA Forest Service. A guide to common insects and diseases of forest trees in the northeastern United States. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Area State and Private Forestry, Forest Insect and Disease Management; 1979b. 123 p.

Valentine, Walter Dewey. Analysis of ecosystem models. Ph.D. thesis. Utah State University. Ann Arbor, MI: University Microfilms International, Microfilm No. 7917990; 1978; 165 p.

Holdaway, Margaret R.; Brand, Gary J.

An evaluation of the STEMS tree growth projection system. Res. Pap. NC-234. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1983. 20 p.

STEMS (Stand and Tree Evaluation and Modeling System) is a tree growth projection system. This paper (1) compares the performance of the current version of STEMS developed for the Lake States with that of the original model and (2) reports the results of an analysis of the current model over a wide range of conditions and identifies its main strengths and weaknesses.

KEY WORDS: Validation, testing, growth model, Lake States species, verification.

☆ U.S. Government Printing Office: 1983—666-751/5034 Region No. 6