Testing the Accuracy of Growth and Yield Models for Southern Hardwood Forests

H. Michael Rauscher, *USDA Forest Service, Southern Research Station, Bent Creek Research and Demonstration Forest, 1577 Brevard Rd., Asheville, NC 28806;* **Michael J. Young,** *Champion International Corporation, Lake States Region Forest Resources, W6582 U.S. Highway* **2,** *Norway, MI* **49870; Charles D. Webb,** *Webb Forest Decisions, 107 Spring Hollow Lane, Cary, NC 27511;* **Daniel J. Rohison,** *Hardwood Research Cooperative, North Carolina State University, Box 8008, Raleigh, NC 27695-8008.*

ABSTRACT: *The accuracy of ten growth and yield models for Southern Appalachian upland hardwood forests and southern bottomlandforests was evaluated. In technical applications, accuracy is the composite both bias (average error) andprecision. Results indicate that GHAT, NATPIS, and a locally calibrated version of NETWIGS may be regarded as being operationally valid growth and yield models for Southern Appalachian yellow-poplar* (Liriodendron tulipifera) *and mixed oak* (Quercus *spp.) forests that fall within the range of characteristics of the test data set. No publicly available growth and yield models specifically developed for southern bottomland hardwoodforests exist, Four general models that contain most of the applicable species to predict growth of these forests were tested. SETWIGS was found to be the most accurate of the four models tested and is recommended for use if the reported level of accuracy is acceptable and the target stand characteristics fall within the range of our test data set. Results indicate that the growth and density dynamics of dense, young stands of both upland and bottomland hardwoods were poorly predicted by the models. Models predicted basal area and density changes in yellow-poplar stands more accurately than mixed hardwoods. Predictions for upland hardwoods were more accurate than those for bottomland hardwoods. Model accuracy uniformly decreases with increasing length of the projection period. South. J. Appl. For. 24(3):176–185.*

Because the accuracy of growth and yield model projections is likely to affect the quality of forest management decisions, models should be evaluated in order to build confidence in their validity (Vanclay et al. 1996). Rykiel(1996) identified three types of model evaluation: (1) operational validation, (2) conceptual validation, and (3) data validation. Operational validation tests whether the model output conforms with its stated purpose. Conceptual validation evaluates whether the model provides a scientifically acceptable explanation of the causal relationships included in the model. Finally, data validation checks whether the test data accurately represent the system of interest. Forest growth and yield models, unlike many ecological simulation models, are primarily intended for use by forest managers in the decision-making process. Therefore, operational validation is of primary importance and will be considered in this article. Conceptual validation is, in this context, a secondary consideration and is not addressed here. The validity of data is particularly important in the context of model evaluation.

The operational validation process for growth and yield models may be viewed from two perspectives: (1) hypothesis testing and (2) confidence interval estimation (Reynolds 1984, Rykiel 1996). Hypothesis testing procedures are the appropriate tools to determine whether the model mimics the real world by passing some predefined level of accuracy. Confidence interval estimation establishes the degree of confidence that can be placed in model predictions through estimates of the direction, magnitude, and variability of the prediction error (Reynolds 1984). Because generally accepted standards of accuracy for growth and yield models do not exist and vary from one situation to another, the confidence interval perspective to operational validation is fre-

NOTE: H. Michael Rauscher is the corresponding author, and he can be reached at (828) 667-5261 ext. 102; E-mail: mrauscher@fs.fed.us. Long-term forest research plots tend to have a high front end cost with little immediate reward to individuals who design the study and establish the plots. Benefits are often reaped long after the initiators have retired. This study could not have been done without the vision and hard work of Lino Della-Bianca (retired USDA Forest Service), Donald E. Beck (retired USDA Forest Service), Ralph M. Hooper (retired USDA Forest Service), Robert Kellison (retired NC-State Hardwood Research Cooperative),Virginia Gibbs, Tracy Roof and Julia Murphy, forestry research technician, Bent Creek Experimental, Asheville, NC. The plot network within the NC State Hardwood Research Cooperative could not have been possible without the support, and commitment of its industrial and state agency members throughout the South since 1963. We thank Carole Beam, research statistician with the NC State-Hardwood Research Cooperative, for her hard work in the data analysis phase of this project. Our thanks to the following reviewers: Marilyn Buford, Harold Burkhart, George Gertner, Jeff Goelz, Thomas Lynch, Ralph Meldahl, Richard Oderwald, Edward Rykiel, John Stanturf, Susan Stout, and four anonymous reviewers. Manuscript received September 13, accepted January 3 1, 2000.

quently more practical than the hypothesis testing perspective and is the approach used in this study (Goodall 1972).

Objectives of this article are to test the accuracy of predicting the changes in basal area growth and stand density in: (1) ten growth and yield models for Southern Appalachian upland hardwood forests, and (2) four growth and yield models for southern bottomland hardwood forests.

The Southern Appalachian region covers approximately 37 million ac in parts of seven states: Virginia, North Carolina, Tennessee, Kentucky, South Carolina, Georgia, and Alabama (SAMAB 1996). Forests currently cover 70% of the region. The Southern Appalachian upland hardwood part of this study concentrates on the 16 million ac of the region in either mixed oak (Quercus spp.) or yellow-poplar (*Liriodendron tulipifera*) forest. The southern bottomland hardwood part of this study concentrates on the approximately 23 million ac in the eight Southern states (Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Virginia) wetland hardwood forest types (USDA 1988). Bottomland sites are considered those in which 50% or more of the forest stand is comprised of tupelo *(Nyssa* spp.), sweetgum *(Liquidambar styraciflua*), oak *(Quercus* spp.), and cypress (*Taxodium* spp.), singly or in combination, with pines (*Pinus* spp.) contributing less than 25% of the stand. There is, however, a good deal of discrepancy in how wetland forests are defined and measured (Cubbage and Flather 1993).

Methods

The Models

A literature review revealed ten candidate growth and yield models with the potential for predicting growth in southern hardwood forests (Table 1). To be included, a model had to be publicly available and be able to forecast the growth of the appropriate species found in the region. Four of the models, CSTWIGS, NETWIGS, OAKSIM, and SILVAH, were developed outside the region of interest for this study. Their authors never intended these models to be used in the South. However, the species and stand structures for which these models were developed are similar to those in the southern region, and it seemed reasonable to test their utility in this region. To the extent that these models have been validated before (Schuler et al. 1993), this study may be viewed as a revalidation test to determine whether the model is acceptable for use in a new region (Rykiel 1996). If any of these four were to pass this test, then its applicability would increase.

Two more models, SETWIGS and GATWIGS, were developed from USDA Forest Service, Forest Inventory and Analysis data (Table 1). NATPIS was developed from yield tables directly applicable to the Southern Appalachian region. The remaining three models, GHAT, MIXOAK, and YPOP, were all developed from permanent research plot data within the Southern Appalachian region.

All ten models in Table 1 were evaluated for accuracy using the Southern Appalachian upland hardwood forest data set (hereafter referred to as the "upland hardwood" data set). Only four of the ten models, NETWIGS, CSTWIGS, GATWIGS, and SETWIGS, could be used to simulate bottomland hardwood forest growth because only these contained the required bottomland species. These models were evaluated for accuracy using the southern bottomland hardwood forest data set (hereafter referred to as the "bottomland hardwood" data set).

The Upland Hardwood Test Data Set

Data consisted of 236 permanent plots from 4 ongoing experiments: (1) Yellow-Poplar Stand Density Study (Olson 1959, Della-Bianca 1965); (2) Mixed Hardwood Stand Density Study (Beck 1973); (3) Mixed Hardwood Unthinned Control Plots (Beck 1973); and (4) Yields of Unthinned Mixed Hardwood Stands (Smith et al. 1975). The plots were located in North Carolina, Virginia, Georgia, Tennessee, and South Carolina and were remeasured at 5 yr intervals. All but 32 plots were thinned from below at the time of installation to obtain a range of basal areas for different site-age combinations.

For practical reasons, the test data set had to be reduced in size to be manageable. Many of the models tested could not execute a batch of plots automatically, thereby requiring each observation in the test data set to be individually run. A 10 yr remeasurement interval was used to represent short-term growth predictions. The data yielded 397 such 10 yr intervals without overlapping the measurement year. Cluster analysis, using SAS PROC CLUSTER (SAS 1990) and specifying Ward's minimum-variance method of clustering with a 5% trimming of the data to exclude outlier influence and data standardization to reduce undue influence of variables with large variance, was used to identify homogeneous clusters of

Table 1. Description of the ten growth and yield models included in the accuracy test for Southern Appalachian hardwood forests.

Model name	Author(s)	Scale	Type of data	Origin of data	Test data set used
CSTWIGS	Miner et al. 1988	Individual tree	FIA* data & permanent research plots	IN, OH, MO	Upland & bottomland
GATWIGS	Bolton & Meldahl 1990	Individual tree	FIA data	GA	Upland & bottomland
GHAT	Harrison et al. 1986	Individual tree	Permanent research plots	GA, NC, TN, VA	Upland
MIXOAK	Bowling et al. 1989	Diameter distribution	Permanent research plots	GA, NC, TN, VA	Upland
NATPIS	Smith & Hafley 1987	Stand	Yield table	NC, SC, GA	Upland
NETWIGS	Hilt & Teck 1989	Individual tree	FIA data	Northeastern US	Upland & bottomland
OAKSIM	Hilt 1985	Individual tree	Permanent research plots	OH, KY	Upland
SETWIGS	Bolton & Meldahl 1989	Individual tree	FIA data	AL, GA, SC	Upland & bottomland
SILVAH	Marquis & Ernst 1992	Stand-table projection	Permanent research plots	PA, OH, KY	Upland
YPOP	Knoebel et al. 1986	Diameter distribution	Permanent research plots	GA, NC, VA	Upland

* USDA Forest Service, Forest Inventory and Analysis

observations with one cluster reserved for outliers. Observations were randomly selected from each of the clusters including the cluster of outliers.

A total of 184 observations were used with time intervals of 10 yr as the minimum and 25 yr as the maximum. The test data set had the following characteristics: even-aged, yellow-poplar or mixed oak dominated overstory, dry-mesic to wet-mesic moisture regimes, 20 to 100 yr in age, and 25 to 230 ft^2/ac of basal area. The site index range was 74 to 138 ft (base age 50) for yellow-poplar and 60 to 110 ft (base age 50 for northern red *Q. rubra*) for mixed oak.

The Bottomland Test Data Set

The data set available for evaluating the growth of bottom land hardwoods consisted of 49 permanent plots established by the NC State Hardwood Research Cooperative between 1969 and 1974 (Smith et al. 1975). Plots were grouped by site type as: (1) nonswamp forests (41 plots) and (2) swamp forests (8 plots) (Kellison et al. 1982). Nonswamp forests are generally characterized by mineral soils, relatively rapid drainage, floodplain and alluvial landscape positions, and a wide variety of tree species, including sweetgum (*Liquidambar styraciflua*), ash (***Fraxinus*** spp.), oak (***Quercus*** spp.), and red maple (*Acer rubrum*). Swamp forest sites are characterized by organic soils or heavy accumulations of organic material over mineral soil, relatively slow drainage or flooded conditions, extremely low or depressed landscape positions, and they are dominated by relatively few tree species, such as water tupelo (*Nyssa* ***aquatica),*** black gum (*Nyssa sylvatica*), and bald cypress ***(Taxodium distichum).*** These two categories, while ignoring the great variety of specific site and stand types across the southern bottomland region, provide a useful division in large-scale growth and yield model evaluations, such as in the current study.

Plots were located in reasonably well-stocked stands, avoiding stands showing evidence of major disturbance. No thinnings wereconducted. At establishment, the stands ranged from 10 to 40 yr in age, 66 to 125 ft in site index (base age 25 yr), and 31 to 373 ft^2/ac of basal area. Plots were remeasured as many as 6 times on a 5 yr remeasurement cycle. A total of 107 observations were used from the 49 permanent sample plots with time intervals of 10 yr as the minimum and 25 yr as the maximum.

Performance Measures

In technical parlance, accuracy refers to the size of the deviation between observed and predicted values. Accuracy has two components: bias and precision. Bias refers to the success of estirnating the true value of a quantity and precision refers to the clustering of sample values about their own average. Freese (1960) pointed out problems with using a *t*-test for this purpose. Reynolds (1984) expanded on the work by Freese and presented a complete system for testing accuracy. Software to implement the methodology advocated by Reynolds (1984) has been developed (Rauscher 1986) and further refined (Gribko and Wiant 1993, Wiant 1993). Accuracy testing studies using this method have been widely reported (Brand and Holdaway 1989, Schuler et al. 1993, Wiant et al. 1996).

The DOSATEST program (Wiant 1993) was used to calculate bias in percent and whether the level of bias was significant significantly different from zero. The software also calculated a tolerance interval (TI) which indicates the limit that contains 75% of all future prediction errors 95% of the time. This limit is calculated as the mean bias plus and minus the TI (Wiant et al. 1996).

The mean square error (MSE) and prediction accuracy at ±15% (PA-15) statistics were calculated because they combine the effect of bias and precision to yield a unified estimate of prediction accuracy (Devore'1982, Clutter and Gent 1992). The MSE of the residuals, a commonly calculated statistic, has one major drawback. The MSE of one study cannot be used to compare results with other studies. The PA- 15 statistic is a readily understandable and simple to calculate estimator of the proportion of time the candidate model predictions come within ±15% of the observed value. The PA- 15 statistic is a somewhat arbitrary but reasonable standard suggested by Schuler et al. (1993) and Rykiel(1996). The PA-15 statistic, unlike the MSE, can be used to compare the results of one accuracy testing study to another. In general, the greater the bias and/or the less the precision, the less accurate is the estimator which translates into a larger MSE and a smaller PA- 15 value.

Test Procedures

Each of the ten candidate models, except GHAT, MIXOAK, and YPOP, was first tested against the entire upland hardwood data set of 184 observations. GHAT, MIXOAK, and YPOP were tested against 164 observations of the upland hardwood data set, eliminating data which had been used to fit the coefficients of these models. Four models, CSTWIGS, GATWIGS, NETWIGS, and SETWIGS were tested against the entire bottomland hardwood data set of 107 observations. In both cases, model predictions for stand basal area and stand density (trees per acre) were compared to observed values for each observation in the test data sets and the results averaged into a single MSE and PA- 15 value for each model. This average value was then used to establish the overall accuracy ranking of the models.

Tree growth and mortality are the major components of growth models. Stand basal area was selected to test the growth prediction component of each of the ten models. It is a directly estimated and observable quantity. Similarly, stand density was used to test the mortality component of each of the models. Dbh is redundant because it is functionally related to stand basal area and density. Volume was not used because it is a derived attribute (Brand and Holdaway 1989). Furthermore, volume causes discontinuities due to merchantability criteria that have nothing to do with the physical dynamics of tree growth and mortality.

The top ranked models were chosen to undergo a more detailed analysis of their prediction performance where accuracy was decomposed into separate estimates for bias and precision. The detailed analyses examined various subsets of the data to determine how accuracy was influenced by forest or site types, treatment, site index, basal area, trees per acre, age, and growth interval length comparisons. Forest type comparisons were between mixed oak and yellow-poplar for the upland hardwood data set, and between swamp and nonswamp site types for the bottomland hardwood data set. Treatment comparisons for the upland test data set were thinned from below versus no treatment. The bottomland hardwood test data set contained no thinned plots. Growth interval comparisons were short (≤10 yr) versus long (>10 yr). The other continuous variables-site index, basal area, trees per acre, and age-were subdivided into lower quartile, middle half, and upper quartile subsets.

Results

Overall Model Performance

For upland hardwoods in the Southern Appalachian region, the GHAT and YPOP models were the most accurate of the ten models tested as measured by the average of the PA-15 and MSE values for stand basal area and stand density (Table 2). The fact that YPOP is limited to predicting the growth of yellow-poplar dominated forest stands whereas GHAT is able to predict both yellow-poplar and oak dominated forest stands, leads us to rank GHAT as the more generally useful model. When GHAT is evaluated for yellow-poplar dominated stands only, its performance is as good as that of YPOP (Table 2).

NATPIS and MIXOAK make up the next most accurate group of models in this study (Table 2). NATPIS is interesting because it is the only stand-level model in this evaluation and the only one based on published yield tables instead of plot data. The input requirements for NATPIS, basal area and trees/ac for up to five species groups, are less expensive to acquire than the input requirements for tree level models such as GHAT. MIXOAK, one of two diameter distribution models evaluated, shares the modest input requirements of NATPIS, but is not as accurate as either NATPIS or GHAT.

NETWIGS, SILVAH, and OAKSIM make up the third most accurate group of models in this study (Table 2). NETWIGS and SILVAH are the highest ranking, extra-regional models. On the basis of accuracy (MSE and PA-15) alone, they cannot be separated. NETWIGS may be more attractive to users than SILVAH because it has been imported into the USDA Forest Service's Forest Vegetation Simulator, a widely used, well supported growth and yield model management system (Teck et al. 1996). Although substantially less accurate in this test than either GHAT or NATPIS, NETWIGS is more comprehensive because it can be used to forecast growth for the other 26% of the Southern Appalachian forested landscape that neither of the two more accurate models cover. OAKSIM can only be used for oak dominated forests and is somewhat less accurate than either NETWIGS or SILVAH. Based largely on desirable model features other than accuracy, we selected NETWIGS rather than SILVAH to represent the "best" of the extra-regional models for the more detailed subset comparisons reported in the next section.

The remaining models, GATWIGS, CSTWIGS, and SETWIGS cannot be recommended for predicting upland hardwood growth and yield in the Southern Appalachian region. These models may do very well when restricted to their intended regions and forest types, but they do poorly for the medium to high quality oak and yellow-poplar dominated forests in the Southern Appalachian region.

SETWIGS proved the most accurate of the four models tested for the Southern bottomland hardwood data (Table 2), with a PA-15 value of 71%. NETWIGS, GATWIGS, and CSTWIGS were much less accurate with PA-15 values ranging from 5 1 to 57%.

Specific Model Performance

GHAT, NATPIS, and NETWIGS for the upland hardwoods and SETWIGS for the bottomland hardwoods were subjected to a more detailed analysis in order to identify specific forest conditions where these models do exceptionally well or exceptionally poorly. In general, dense, young stands of both upland

Table 2. Results of the accuracy test for ten growth and yield models using prediction accuracy and mean square error as ranking metrics.

Model name	Rank	PA-15* (%)	MSE† (%)	Number of observations	Remarks
Southern Appalachian hardwoods					
YPOP	—	93	53	99	Yellow-poplar plots only
GHAT††	1	86	105	163	21 partitioned plots removed
NATPIS	2	8.5	125	184	
MIXOAK	3	84	170	163	2 1 partitioned plots removed
NETWIGS	4	78	223	184	
SILVAH	5	75	213	184	
OAKSIM		73	219	85	Mixed hardwood plots only
GATWIGS	6	65	267	184	
CSTWIGS	7	65	298	184	
SETWIGS	8	56	358	184	
Southern bottomland hardwoods					
SETWIGS	1	71	656	107	
NETWIGS	2	57	678	107	
GATWIGS	3	54	726	107	
CSTWIGS	4	51	786	107	

* PA-15 = the average accuracy that the predicted value fell within ±15% of the observed value for basal area and trees per acre estimates combined.

† MSE = the average mean square error for basal area and trees per acre estimates combined on a percent basis.

†† For yellow-poplar test data only, GHAT results are: PA-15 = 94%; MSE = 55%.

and bottomland hardwoods are poorly predicted by the models (Tables 3-5). The youngest plots in the test data set were 18 yr old for the upland hardwood data set and 10 yr old for the bottomland hardwood data set.

GHAT

GHAT was found to be an unbiased predictor of yellow-poplar and mixed oak basal area growth and density when tested against the full test data set (Table 3). The model is capable of predicting future values of basal area to within 15% of the observed values and trees per acre to within 12% of the observed values 75% of the time with 95% probability. Different levels of performance are found as the test data set is subdivided by species, treatment, site index, stocking density, age, and growth-interval length (‘Table 3). Yellow-poplar plots were more accurately predicted than mixed oak plots. GHAT underpredicts mixed oak basal area growth by 4%, and the predictions are much less precise than those for yellow-poplar basal area growth. The tolerance interval is 5% greater than for yellow-poplar and the PA- 15 value decreases from 91 to 72%. GHAT underpredicts yellow-poplar density by 1%. After being thinned, yellow-poplar plots had less mortality than GHAT predicts. Stand density predictions are also more accurate for yellow-poplar than they are for mixed hardwood with GHAT. It seems yellow-poplar basal area growth and density changes are easier to predict than the same values for mixed oak. This result is not unexpected given the much greater degree of variability in mixed oak stands.

For basal area growth predictions, the following results can be observed from Table 3:

1. thinned plots were underpredicted by 2%, while plots with no treatment exhibited greater variability and consequently lower accuracy;
2. plots in the middle half of the site index range were well predicted, upper quartile site index plots had greater variability, and lower quartile site index plots were underpredicted by 5% as well as exhibiting lower precision;
3. plots in the lower quartile of the basal area stocking range were less precisely estimated;
4. plots in the upper quartile of the density stocking range were underpredicted by 4%;
5. plots with younger trees were predicted less precisely; and
6. the short-growth interval plots were underpredicted by 5%, whereas the long-growth interval plots were overpredicted by 3% with an associated decrease in precision.

Table 3. A detailed error analysis for GHAT and NATPIS for Southern Appalachian hardwoods using basal area (BA) and trees per acre (TPA).

		GHAT								NATPIS			
Test description		BA bias*	BA TI†	BA PA-15††	TPA bias	TPA TI	TPA PA-15	BA bias	BA TI	BA PA-15	TPA bias	TPA TI	TPA PA-15
							(%)						
Species	YP	0	12	91	-1	5	98	-3	12	89	-2	7	97
	MO	–4	17	72	0	17	83	-6	19	69	3	17	83
Treatment	‘Thinned	-2	14	85	-1	10	95	-3	14	84	0	12	93
	-None	0	19	69	9	15	77	-8	20	64	0	17	82
Site index	LQ	-5	17	76	0	11	93	-6	15	78	0	12	91
	MH	0	13	87	0	13	92	-5	15	85	0	14	91
	UQ	0	18	79	0	14	88	0	19	72	0	13	91
Basal area	LQ	0	19	77	0	12	93	0	17	80	0	13	95
	MH	0	14	82	0	12	91	-6	14	83	0	14	86
	UQ	0	13	90	0	14	90	-9	16	72	0	13	95
TPA	LQ	0	16	79	0	8	96	0	15	83	0	9	96
	MH	0	16	82	0	13	91	–4	15	82	0	14	90
	UQ	–4	15	85	0	15	86	-9	17	73	0	15	86
Age	LQ	0	18	79	0	15	90	0	16	82	0	16	90
	MH	0	14	83	0	11	92	-6	17	73	0	10	94
	UQ	0	15	83	0	13	90	–4	12	88	0	14	86
Growth interval	Short	-5	12	88	0	10	95	-3	13	83	0	10	96
	Long	3	17	12	0	15	85	-5	19	75	0	17	83
All test data comparison		0	15	82	0	12	91	–4	15	80	0	13	91

* A positive bias means model over-prediction; a negative bias means model underprediction.
† TI = tolerance interval, which is the limit that contains 75% of all future observations 95% of the time.
†† PA-15 = percent of prediction errors falling within ±15% of observed values. Abbreviations are: YP (Yellow Poplar), MO (Mixed Oak), LQ (Lower Quartile), MH (Middle Half), UQ (Upper Quartile).

For stand density predictions, the following was observed from Table 3:

1. on thinned plots, density was underpredicted by 1% but precision was high;
2. on unthinned plots, density was overpredicted by 9% accompanied by lower precision; and
3. density was estimated with noteworthy accuracy (no bias and high precision) in most other tests.

NATPIS

NATPIS was found to be a biased estimator for predicted basal area growth and an unbiased predictor of stand density (Table 3). As pointed out previously for GHAT, NATPIS also illustrates that yellow-poplar basal area, and density can be predicted more precisely than the same variables for mixed oak, reflecting the higher degree of variation in mixed oak stands. NATPIS predicts basal area growth very well for young stands with low basal area and low density. It predicts stand density well for all combinations of basal area and trees per acre. Basal area growth predictions are less accurate for middle aged stands with no treatment, for average to low site index stands, and for stands with average to high basal area and density. Predictions for unthinned stands are less accurate than those for thinned stands.

NETWIGS

NETWIGS was found to be a biased estimator for both basal area growth (-9%) and trees per acre (-8%) (Table 4). The fact that precision, as measured by the TI, is not much worse than that for GHAT or NATPIS, makes it possible to correct for bias. Predicted basal area for each plot was increased by 9%. Similarly, predicted trees per acre for each plot was increased by 8%. The results of this "locally calibrated" evaluation are in Table 4. After correction for bias, the locally calibrated version of NETWIGS was found to be an unbiased estimator for both basal area growth and trees per acre. The model underpredicts high basal area stands by 4%. It is less precise for mixed hardwood stands, for stands of average basal area, for low density stands, and for older stands. NETWIGS also underpredicts yellow-poplar density by 2% (calculates higher mortality than observed on the plots), overpredicts density of no-treatment stands by 5% (calculates lower mortality than observed), underpredicts density for high quality sites by 4%, and underpredicts density for long growth intervals by 4%. Again, predictions for unthinned stands are less accurate than those for thinned stands.

Table 4. A detailed error analysisfor NETWIGS and the locally calibrated version of NETWIGS for Southern Appalachian hardwoodsusing basal area (BA) and trees per acre (TPA).

Test description		NETWIGS						NETWIGS (locally calibrated)					
		BA bias*	BA TI†	BA PA-15††	TPA bias	TPA TI	TPA PA-15	BA bias	BA TI	BA PA-15	TPA bias	TPA TI	TPA PA-15
		(%)											
Species	YP	-9	13	75	-9	9	82	0	14	86	–2	10	93
	MO	-8	22	71	-5	20	87	0	24	76	0	22	83
Treatment	Thinned	-9	17	74	-8	16	83	0	19	85	0	18	90
	None	-5	19	69	0	13	90	0	21	67	5	14	82
Site index	LQ	-7	15	74	-6	10	96	0	17	85	0	11	93
	MH	-8	20	80	-7	20	85	0	21	84	0	21	88
	UQ	-11	18	57	-11	14	68	0	20	72	–4	15	84
Basal area	LQ	-9	14	71	-7	11	89	0	16	87	0	12	93
	MH	-7	21	79	-7	20	84	0	23	76	0	22	86
	UQ	-12	12	61	-9	13	77	–4	14	85	0	14	87
TPA	LQ	-12	25	66	-7	24	87	0	28	85	0	26	96
	MH	-8	15	76	-7	13	83	0	16	83	0	14	90
	UQ	-6	16	75	-7	15	83	0	18	75	0	16	79
Age	LQ	-6	16	71	-8	15	79	0	17	81	0	17	86
	MH	-10	14	71	-8	10	86	0	15	81	0	11	91
	UQ	-8	28	79	-6	28	88	0	31	83	0	30	88
Growth interval	Short	-7	18	81	-6	16	96	0	20	85	0	18	94
	Long	-11	15	59	-11	15	65	0	17	75	–4	16	80
All test data comparison		-9	17	73	-8	15	84	0	19	81	0	17	89

* A positive bias means model over-prediction; a negative bias means model under-prediction.
† TI = tolerance interval, which is the limit that contains 75% of all future observations 95% of the time.
†† PA-15 = percent of prediction errors falling within +/- 15% of observed values. Abbreviations are: YP (Yellow Poplar), MO (Mixed Oak), LQ (Lower Quartile), MH (Middle Half), UQ (Upper Quartile).

It should be noted that forest managers can correct basal area and density estimates from NETWIGS in the same way we did. The first step is to use NETWIGS to estimate stand basal area and density. Next, increase the predicted basal area by 9% and increase the predicted density value by 8%. The resulting estimates are locally calibrated and are expected to attain the levels of accuracy shown in Table 4.

SETWIGS

SETWIGS was found to be a biased estimator of southern bottomland hardwood growth for both basal area and density when tested against the full bottomland data set (Table 5). The model underestimated basal area by 9% and trees per acre by 6%. SETWIGS was found capable of predicting future values of basal area to within 30% of observed values and density to within 36% of observed values 75% of the time with 95% probability. SETWIGS predicts southern bottomland hardwood forest growth and mortality with less accuracy than GHAT or even NETWIGS does for Appalachian upland hardwood forests.

Different levels of performance were found as the test data set was subdivided by site type, site index, basal area, density, age, and growth interval (Table 5). For basal area growth predictions, the following results were observed for SETWIGS:

1. nonswamp stands can be predicted with less bias than swamp stands;
2. stands with basal area in the lower quartile showed more bias and generally less precision than middle half or upper quartile stands;
3. stands with density in the upper quartile showed more bias and generally less precision than middle half or upper quartile stands;
4. predictions for young stands are substantially poorer than for older stands; and
5. no differences were detected due to length of the growth interval.

For density predictions, the following results were observed for SETWIGS:

1. nonswamp stands can be predicted with less bias than swamp stands;
2. plots with site indices in the lower quartile were much more accurately predicted than any other types of plots;
3. plots with low basal area and high density are very poorly predicted; and
4. trees per acre were overpredicted for older plots and underpredicted for younger plots, indicating that NETWIGS calculates too little and too much mortality for older and younger plot types, respectively.

Table 5. A detailed error analysis for SETWIGS for southern bottomland hardwoods using basal area (BA) and trees per acre (TPA).

Test description		BA bias*	BA TI†	BA PA-15††	TPA bias	TPA TI	TPA PA-15
				(%)			
Site type	Nonswamp	-8	31	73	-4	37	73
	Swamp	-16	28	64	-16	32	57
Site index	Upper quartile	-8	52	71	-7	57	74
	Middle half	-10	17	69	-6	30	61
	Lower quartile	-6	26	78	-1	29	87
BA	Upper quartile	-5	13	88	-5	15	92
	Middle half	-9	17	70	-2	27	70
	Lower quartile	-14	65	59	-16	71	52
TPA	Upper quartile	-12	36	67	-14	36	62
	Middle half	-7	19	79	-5	25	77
	Lower quartile	-8	53	63	6	58	67
Age	Upper quartile	-3	17	82	11	34	73
	Middle half	-7	30	81	-4	34	80
	Lower quartile	-26	41	13	-26	55	20
Growth interval	Short	-8	32	69	-8	36	76
	Long	-9	34	75	-3	42	66
All test data comparison		-9	30	72	-6	36	71

* A positive bias means model over-prediction; a negative bias means model underprediction.
† TI = tolerance interval, which is the limit that contains 75% of all future observations 95% of the time.
†† PA-15 = percent of prediction errors falling within ±15% of observed values.

Discussion

GHAT was able to predict basal area growth and density for Southern Appalachian upland hardwood forests to within ±15% of the observed value 86% of the time. This level of accuracy is surprising given the fact that GHAT was developed using only the first two 5 yr remeasurements of the mixed oak study, one out of the four used in this evaluation. These plots made up only 11% of the total Appalachian test data set and were removed from the data set for the GHAT evaluation. The remaining records contained plots representing different conditions than those used to fit the GHAT model originally. For example, growth intervals of 15-25 yr, unthinned conditions, and plots dominated by yellow-poplar were in the test data set but were not used to develop GHAT (Harrison et al. 1986). This suggests that the growth and noncatastrophic mortality of mature Southern Appalachian hardwood forests may be readily predictable. Our results seem to indicate that it may take only a relatively short period of time, in this case 10 yr, to accumulate enough data to develop usefully accurate predictive models of overstory growth and mortality.

Stands where yellow-poplar dominates the overstory are more accurately predicted than mixed oak stands. GHAT comes to within ±15% of the observed basal areas 91% of the time and to within ±15% of density observations 98% of the time. This result is interesting because 54% of the plots in the test data set were composed of yellow-poplar dominated plots, where this species made up over 90% of the basal area. No such plots were used to develop the GHAT model. These results underscore the intuitive conclusion that the greater the species homogeneity of the overstory, the more accurate forest growth and mortality predictions can be.

GHAT's performance may have been overestimated because the test data was drawn from studies that were used to develop the model coefficients, despite our removal of the specific plots in question for the accuracy test (Tukey 1977). It is therefore interesting to examine NETWIGS, the best of the models that were completely independent of the test data set, in more detail (Table 4). Our study found that NETWIGS can forecast basal area growth and density to within ±15% of the observed value, on average, 78% of the time based on 184 observations. In comparison, Schuler et al. (1993), found that NETWIGS can come within ±15% of the observed value 64% of the time for all stand variables tested for the oak-hickory forest type in New York, Kentucky, and Ohio based on 29 observations. For Ohio alone, NETWIGS can predict to within ±15% of the observed value 79% of the time for all stand variables tested for the oak-hickory forest type, a total of seven observations. The performance of NETWIGS for the oak-hickory forest type in Ohio appears to be equal in accuracy to our results but lower for the northeast region as a whole. These results may be an anomaly of the small sample size of the Schuler et al. (1993) study. NETWIGS underestimates basal area by 9% in the Southern Appalachian region (Table 4) and by 7% in the northeast (Schuler et al. 1993). The model underestimates density by 8% in the Southern Appalachian region (Table 4) and by 4% in the northeast. The TI values are roughly the same in both studies. A bias correction of 9% for basal area and 8% for density can be applied to NETWIGS to improve its accuracy to come within ±15% of the observed basal area and density 85% of the time, on average. This compares favorably with the accuracy test results for GHAT. In this case, a correction of the bias is helpful because the precision, as measured by the TI, is only slightly larger than that of GHAT and does not change substantially as a result of the bias correction.

Our study also confirmed the findings of Holdaway and Brand (1986), Kowalski and Gertner (1989), and Shortt and Burkhart (1996) that accuracy decreases with increasing length of the projection period. Kangas (1997) also found that both variance and bias of predictions increase with increasing simulation time from 5-50 yr. In our study, all test results for GHAT, NATPIS, and NETWIGS corroborated this relationship. Various accuracy testing studies have been able to directly test projection lengths of between 10 and 40 yr. In this study, an increase of projection length from 10 to 20-25 yr using the model GHAT resulted in a reduction in PA-15 from 91% to 78%. Extrapolating this rate of 13% reduction in accuracy for every 20 yr increase in projection length might result in the following relationship:

Projection length (yr):	10	30	50	70	90
PA-IS:	91	78	6.5	52	39

Forest management planning in the USDA Forest Service routinely requires using growth and yield models with projection lengths of 100 to 300 yr. Depending on how results are used, such long projection lengths for growth and yield models may admit unacceptably large inaccuracies into the decision making process. Use of the best models in this evaluation for long-term (e.g. >40 yr) predictions of overstory growth and mortality should be approached with caution. An exponential increase in prediction error levels "far in excess of reasonable model utility" should be expected with projection lengths greater than 30-40 yr (Mowrer and Frayer 1986, Kangas 1997). On the other hand, long projection lengths may still be useful if the goal is to assess overall trends for a timber type to support forest level decision making.

The results of this study indicate that currently available models do not perform as well for southern bottomland hardwood forest conditions as they do for Southern Appalachian upland forest conditions. No specific growth and yield models have yet been developed for southern bottomland hardwoods, so this result is not surprising. However, models developed for other regions and forest types are being used or considered for use with bottomland hardwoods, and we felt it was important to assess their accuracy characteristics to provide guidance for forest management decision-making in southern bottomland forests. Our accuracy evaluation using the bottomland test data set resulted in SETWIGS as the most accurate model of the four tested (Table 2). SETWIGS was able to predict basal area and trees per acre to within ±15% of the observed values 71% of the time, on average, based on 107 observations. The detailed test comparisons showed that SETWIGS is a biased predictor (Table 4), underpredicting basal area by 9% and trees per acre by 6%. For predicting Southern bottomland hardwood dynamics, SETWIGS is also substantially less precise, as measured by the tolerance interval, than GHAT for Southern Appalachian upland hardwoods. A simple bias correction, as suggested for NETWIGS above, is not likely to improve overall accuracy significantly because of poor precision of the estimates. The need for an unbiased and more precise growth and yield model for southern bottomland hardwood conditions is evident.

Conclusion

The accuracy of ten growth and yield models for Southern Appalachian upland hardwood forests and southern bottomland forests was evaluated. GHAT, NATPIS, and the locally calibrated version of NETWIGS may be regarded as being operationally valid models for growth and yield predictions for Southern Appalachian yellow-poplar and mixed oak forests that fall within the range of characteristics of the test data set. In brief, these characteristics are: even-aged, yellow-poplar or mixed oak dominated overstory, dry-mesic to wet-mesic moisture regimes, 20 to 100 yr in age, and 25 to 230 ft^2/ac of basal area. For yellow-poplar dominated stands, the appropriate site index range is 74 to 138 ft (base age 50) and for mixed oak the appropriate site index range is 60 to 110 ft (base age 50 for northern red oak). These results are not valid for: xeric oak forests, mixed-aged hardwood forests, heavily disturbed forests (e.g., those with high-grading), or so-called old growth forest conditions where overstory trees are approaching biological old age and experiencing significant dieback and top breakage.

No publicly available growth and yield model has been specifically developed for southern bottomland hardwood forests. Four general models that contain most of the applicable species to predict growth of these forests were evaluated. SETWIGS was found to be the most accurate of the tested models and is recommended for use if the reported level of accuracy is acceptable and the target stand characteristics fall within the range of our test dataset. These characteristics are: even-aged, swamp or nonswamp sites, 10 to 60 yr in age, and 30 to 370 ft^2/ac of basal area. The appropriate site index range is 66 to 125 ft (base age 25 yr). SETWIGS is, however, far less accurate in predicting stand basal area and density than the best of the comparable upland forest growth and yield models in this study. The need for an unbiased and more precise growth and yield model for southern bottomland hardwood conditions is evident.

Literature Cited

BECK, D.E. 1973. Timber growth and understory development in thinned Appalachian mixed hardwoods. Study plan and supplement 1, USDA For. Serv., FS-SE-1102-14(l) [BC-45], Bent Creek Experimental and Demonstration Forest, Asheville, NC. 22 p. + append.

BOLTON, R.K., AND R.S. MELDAHL. 1989. Summary of modeling efforts for southeastern TWIGS. Internal Rep., School of For., Auburn University. P. 216.

BOLTON, R.K., AND R.S. MELDAHL. 1990. User's guide to a multipurpose forest projection system for southern forests. Alabama Agric. Exp. Sta., Bull. 604. 15 p.

BOWLING, E.H., H.E. BURKHART, T.E. BURK, AND D.E. BECK. 1989. A stand level growth model for Appalachian hardwoods. Can. J. For. Res. 19:405-412.

BRAND, G.J., AND M.R. HOLDAWAY. 1989. Assessing the accuracy of TWIGS and STEMS85 volume predictions: A new approach. North. J. Appl. For. 6: 109-l 14.

CLUTTER, M.L., AND J.A. GENT. 1992. Validation and comparison of four cut-over site prepared loblolly pine growth and yield models. P. 593-600 in Proc. of the Seventh South. Silvic. Res. Conf., Brissette, J.C. (ed.). USDA For. Serv. Gen. Tech. Rep. SO-93.

CUBBAGE, F.W., AND C.H. FLATHER. 1993. Forested wetland area and distribution. J. For. 91:35–40.

DELLA-BIANCA, L. 1965. A study of the effects of stand density upon growth and yield of pure yellow-poplar in the Southern Appalachian mountains. Establishmentreportandprogressreport. Southeast. For. Exp. Sta., Study No. M-1.03, FS-SE-1102. 19 p.

DEVORE, J.L. 1982. Probability and statistics for engineering and the sciences. Brooks Cole Publ. Co., Monterey, CA. 640 p.

FREESE, F. 1960. Testing accuracy. For. Sci. 6:139–145.

GOODALL, W.W. 1972. Building and testing ecosystem models. P. 173-194 Mathematical models in ecology, Jeffers, J.N.J. (ed.). Blackwell, Oxford.

GRIBKO, L.S., AND H.V. WIANT, JR. 1993. A SAS template program for the accuracy test. Compiler 1 0(1):48-5 1.

HARRISON, WC., H.E. BURKHART, T.E. BURK, AND D.E. BECK. 1986. Growth and yield of Appalachian mixed hardwoods after thinning. VPI&SU, Sch. For. Wildl. Res., Publ. No. FWS-1-86.48 p.

HILT, D.E. 1985. OAKSIM: An individual-tree growth and yield simulator for managed, even-aged upland oak stands. USDA For. Serv. Res. Pap. NE-562.21 p.

HILT, D.E., AND R.M. TECK. 1989. NE-TWIGS: Anindividual-treegrowthand yield projection system for the northeastern United States. Compiler 7(2):10-16.

HOLDAWAY, M.R., AND G.J. BRAND. 1986. An evaluation of Lake States STEMS85. USDA For. Serv. Res. Pap. NC-269. 10 p.

KANGAS, A.S. 1997. On the prediction bias and variance in long-term growth projections. For. Ecol. Manage. 96:207–216.

KELLISON, R.C., D.J. FREDERICK, AND W.E. GARDNER. 1982. A guide for regenerating and managing natural stands of southern hardwoods. North Carolina State Univ., Agric. Res. Serv. Bull. 463, Raleigh, NC. 23 p.

KNOEBEL, B.R., H.E. BURKHART, AND D.E. BECK. 1986. A growth and yield model for thinned stands of yellow-poplar. For. Sci. Monogr. 27.62 p.

KOWALSKI, D.G., AND G.Z. GERTNER. 1989. A validation of TWIGS for Illinois forests. North. J. Appl. For. 6:154–156.

MARQUIS, D.A., AND R. ERNST. 1992. User'sguide to SILVAH: Standanalysis, prescription, and management simulator program for hardwood stands of the Alleghenies. USDA For. Serv. Gen. Tech. Rep. NE-162. 124 p.

MINER, CL., N.R. WALTERS, AND M.L. BELLI. 1988. A guide to the TWIGS program for the North Central United States. USDA For. Serv. Gen. Tech. Rep. NC-125. 105 p.

MOWRER, H.T., AND W.E. FRAYER. 1986. Variance propagation in growth and yield projections. Can. J. For. Res. 16: 1196–1200.

OLSON, D.F., JR. 1959. A study of the effect of stand density upon growth and yield of pure yellow poplar in the Southern Appalachian Mountains. Working Plan No. FS-1-f7-2, Study No. M-1.03. Southeast. For. Exu. Sta., Asheville, NC. 21 p.

RAUSCHER, H.M. 1986. The microcomputer scientific software series 4: Testing prediction accuracy. USDA For. Serv. Gen. Tech. Rep. NC-107. 19 p.

REYNOLDS, M.R. 1984. Estimating the error in model predictions. For. Sci. 30:454-469.

RYKIEL, E.J., JR. 1996. Testing ecological models: the meaning of validation. Ecol. Model. 90:229–244.

SAMAB. 1996. The Southern Appalachian assessment. Prepared by Federal and State Agencies coordinated through the Southern Appalachian Man and the Biosphere Cooperative, Reports l-5.

SAS. 1990. SAS/STAT User's Guide Volume 1, ACECLUS-FREQ. Version 6, Fourth Edition. SAS Inst. Inc., Cary, NC. 890 p.

SCHULER, T.M., D.A. MARQUIS, R.L. ERNST, AND B.T. SIMPSON. 1993. Test of four growth simulators for the Northeastern United States. USDA For. Serv. Res. Pap. NE-676. 14 p.

SHORTT, J.S., AND H.E. BURKHART. 1996. A comparison of loblolly pine plantation growth and yield models for inventory updating. South. J. Appl. For. 20:15–22.

SMITH, H.D., W.L. HAFLEY, D.L. HOLLEY, AND R.C. KELLISON. 1975. Yields of mixed hardwood stands occurring naturally on a variety of sties in the Southern United States. North Carolina State Univ., School For. Res., Raleigh, NC., Tech. Rep. No. 55.32 p.

SMITH, W.D., AND W.L. HAFLEY. 1987. Simulating the effect of hardwood encroachment on loblolly pine plantations. P. 180-186 *in* Proc. Fourth Bienn. South. Silvic. Res. Conf., USDA For. Serv. Gen. Tech. Rep. SE-42.

TUKEY, J.W. 1977. Exploratory data analysis. Addison-Wesley, Reading, MA. 688 p.

USDA Forest Service. 1988. The South's Fourth Forest: Alternatives for the Future. USDA For. Serv. For. Res. Rep. No. 24. Washington, DC. 512 p.

VANCLAY, J.K., J.P. SKOVSGAARD, AND 0. GARCIA. 1996. Evaluating forest growth models. P. 11-22 in Proc. of. IUFRO S4.11-00, XX World Congress, Gertner, M.K. and G.Z. (eds.), Swiss Federal Inst. for Forest, Snow, and Landscape Research, Birmensdorf, Switzerland.

WIANT, H.V., JR. 1993. DOSATEST for testing accuracy. Compiler 11(3):28–29.

WIANT, H.V., JR., G.B. WOOD, AND M. WILLIAMS. 1996. Comparison of three modem methods for estimating volume of sample trees using one or two diameter measurements. For. Ecol. Manage. 83: 13-16.