Comparing the STEMS and AFIS Growth Models With Respect to the Uncertainty of Predictions

Ronald E. McRoberts, Margaret R. Holdaway, and Veronica C. Lessard

Abstract.—The uncertainty in 5-, 10-, and 20-year diameter growth predictions is estimated using Monte Carlo simulations for four Lake States tree species. Two sets of diameter growth models are used: recalibrations of the STEMS models using forest inventory and analysis data, and new growth models developed as a component of an annual forest inventory system for the North Central region of the United States. Particular attention is focused on the effects that uncertainty in the values of predictor variables has on the uncertainty of model predictions. The cumulative effects of uncertainty in model predictions for individual trees on the uncertainty of basal area predictions at the plot and regional levels are also investigated.

The North Central Research Station (NCRS) uses the STEMS (Belcher *et al.* 1982) diameter growth models as a component of its periodic inventory system to update the status of trees on Forest Inventory and Analysis (FIA) plots not measured in the current cycle. These models predict annual diameter growth for individual trees for pure or mixed and even- or uneven-aged stands for nearly all species in the North Central region of the United States; these models have also been recalibrated for other regions. NCRS is also developing new diameter growth models to be used as a component of an annual forest inventory system. The criterion for selection of predictor variables for these kinds of models is usually a measure of quality of fit of the model to calibration data. However, when non-negligible uncertainty is associated with the values of the predictor variables, then additional criteria ought to be considered. Uncertainty in a predictor variable occurs when large measurement errors are associated with values of the variable, as is frequently the case with ocular measurements, or when the values of the variable are estimates based on samples. When such predictor variables are included in models, the effects of their uncertainties may substantially increase the total uncertainty in model predictions. In statistics, the estimation of parameters under these conditions falls into the category of problems referred to as errors in variables (Fuller 1987).

The objectives of the study, of which this paper is a preliminary report, are threefold: (1) to estimate the total uncertainty in diameter growth model predictions; (2) to investigate the effects of including predictor variables with non-negligible uncertainty in growth models; and (3) to compare the uncertainty of predictions obtained from the STEMS models, recalibrated using Forest Inventory and Analysis (FIA) data, with the uncertainty of predictions obtained from the new models developed as a component of the annual forest inventory system.

Mathematical Statistician, Mathematical Statistician, and Research Forester, respectively, North Central Research Station, USDA Forest Service, 1992 Folwell Avenue, St. Paul, MN, USA.

METHODS

Uncertainty Estimation

Our approach to estimating the uncertainty in model predictions uses Monte Carlo simulations. The essence of the simulation process, explained in detail below, is to initialize plot and tree conditions using actual measurements from permanent FIA plots, add random variation where appropriate to mimic uncertainty, use the models to predict annual diameter growth, record estimates at fixed intervals, and repeat the process many times. As part of the process, plot-level estimates of basal area per acre are obtained by summing results over all trees on plots, and regional estimates of basal area and basal area growth are obtained by summing results over all plots. Because the regional basal area and basal area growth estimates are based on results from only 20 plots (see DATA section), they do not necessarily accurately represent true regional growth estimates. However, these estimates permit investigation of the effects on uncertainty when plot-level estimates are aggregated to produce estimates for larger areas. Finally, although mortality is an important component of growth, it is being addressed in other work, and no provision is made in this study for mortality during the prediction interval.

For this application, a Monte Carlo process is used to generate uncertainty for a variety of sources in two broad categories: first, calibration of the growth models, and second, initial conditions for trees and plots for which the model is used to predict growth. In the first category, uncertainty arises from two sources: (1) residual variation around estimated diameter growth curves; and (2)

uncertainty in the estimates of the growth model parameters. When uncertainty in the predictor variables is negligible and when the models are applied to initial conditions that are known with only negligible uncertainty, then these are the only two sources of variation that need be considered in estimating model prediction uncertainty. For this study, uncertainty for the latter source is excluded, because estimates of parameter covariances for nonlinear models are notoriously unreliable when obtained using analytical methods and because the alternative empirical methods require extensive, time-consuming simulations that have not been completed. When uncertainty in initial tree and plot conditions is non-negligible, then it must be propagated through the models and frequently results in substantially greater model prediction uncertainty. Thus, the Monte Carlo process is also used to generate uncertainty in four additional sources related to the uncertainty of the initial conditions of trees and plots: (3) diameter measurement error; (4) crown ratio observation error; (5) crown class observation error; and (6) variance in site index estimates. It should be noted that uncertainty in dbh measurements generates additional uncertainty in variables derived from dbh such as stand basal area and stand average diameter. In addition, when sampling is conducted for variable radius plots, dbh measurement error affects tree expansion factors and, hence, estimates of numbers of trees per acre.

Estimates of residual variation were obtained as byproducts of calibrating the models (Lessard 2000, Holdaway 2000). Residual variation is assumed to be normally distributed but heterogeneous with standard deviations related to predicted growth as follows:

$$E[\ln(\sigma)]=\beta_1+\beta_2\ln(\Delta d), \qquad [1]$$

where E(.) represents statistical expectation, Δd is annual diameter growth prediction, σ is the standard deviation of the residuals for predicted diameter growth classes of small widths, and the βs are parameters to be estimated.

Distributions representing the uncertainty in crown class, crown ratio, site index, and diameter measurement error were obtained from the literature. Nichols *et al.* (1991) reported that the repeatability of crown class observations by field crews in the same year is approximately 80 percent with the other 20 percent allocated uniformly to the two adjacent classes. McRoberts *et al.* (1994) reported that observations of crown ratio for the same tree by separate field crews ranged $\pm$ 0.3 from the median; they also provide a mathematical model for estimating diameter measurement error as a function of diameter. Finally, McRoberts (1996) reported that coefficients of variation for site index estimates, assumed to be normally distributed, may be as great as 0.15, depending on the number of trees used in calculating the estimate.

The Monte Carlo process for this application involves six steps: (1) mimic uncertainty in initial plot conditions by adding randomly generated variation to plot-level variables such as site index; (2) mimic uncertainty in initial tree conditions by adding randomly generated variation to tree-level variables such as crown ratio, crown class, and diameter; (3) use the models to predict annual diameter growth for each tree, mimic residual uncertainty in the predictions with randomly generated variation, and obtain predicted diameter at the beginning of the subsequent year as the sum of diameter at the beginning of the year, predicted annual diameter growth, and residual variation; (4) repeat step 3 to mimic growth of trees for 20 consecutive years; (5) record diameter predictions for trees, basal area per acre predictions for plots, and basal area and basal area growth predictions for the region at the 5-, 10-, and 20-year intervals; and (6) replicate steps 1-5 many times. In step 5, the basal area per acre estimate for each plot is obtained by multiplying the tree basal area estimate and the corresponding tree expansion factor for all trees and then summing these products for all trees on the plot. Relative estimates of basal area and basal area growth at the regional level are obtained by adding basal area per acre and basal area growth per acre, respectively, over all plots under the simplifying assumptions that each plot represents an area of 1 acre and that the plots represent a simple random sample of the population.

Distributions of the coefficients of variation ($CV=\sigma/\mu$) obtained from 1,000 replications of the simulation process serve as the basis for analyzing the uncertainty of model predictions. Mean values, standard deviations, and coefficients of variation for simulated dbh for individual trees stabilized by completion of 1,000 simulations (fig. 1). The results of the simulations are summarized as median coefficients of variation for diameter estimates of individual trees (table 1), median coefficients of variation for plot-level estimate of basal area (table 2), and regional estimates of basal area and basal area growth (table 3) at 5-, 10-, and 20-year intervals. For individual trees and plot estimates, the summaries are in terms of medians because of possible asymmetry in the distributions.

The uncertainty of model predictions is analyzed with respect to several factors and conditions. First, prediction uncertainty is obtained by conducting the simulations using all sources of variation and serves as a basis for comparing the recalibrated STEMS models and the new models. The contributions to model prediction uncertainty due to uncertainty in individual predictor variables is determined by including the variable when calibrating the model and then comparing the corresponding uncertainties of model predictions when the variable is both excluded and included as a source of uncertainty in the simulations. The results are used to assess the sensitivity of the uncertainty in model predictions to individual

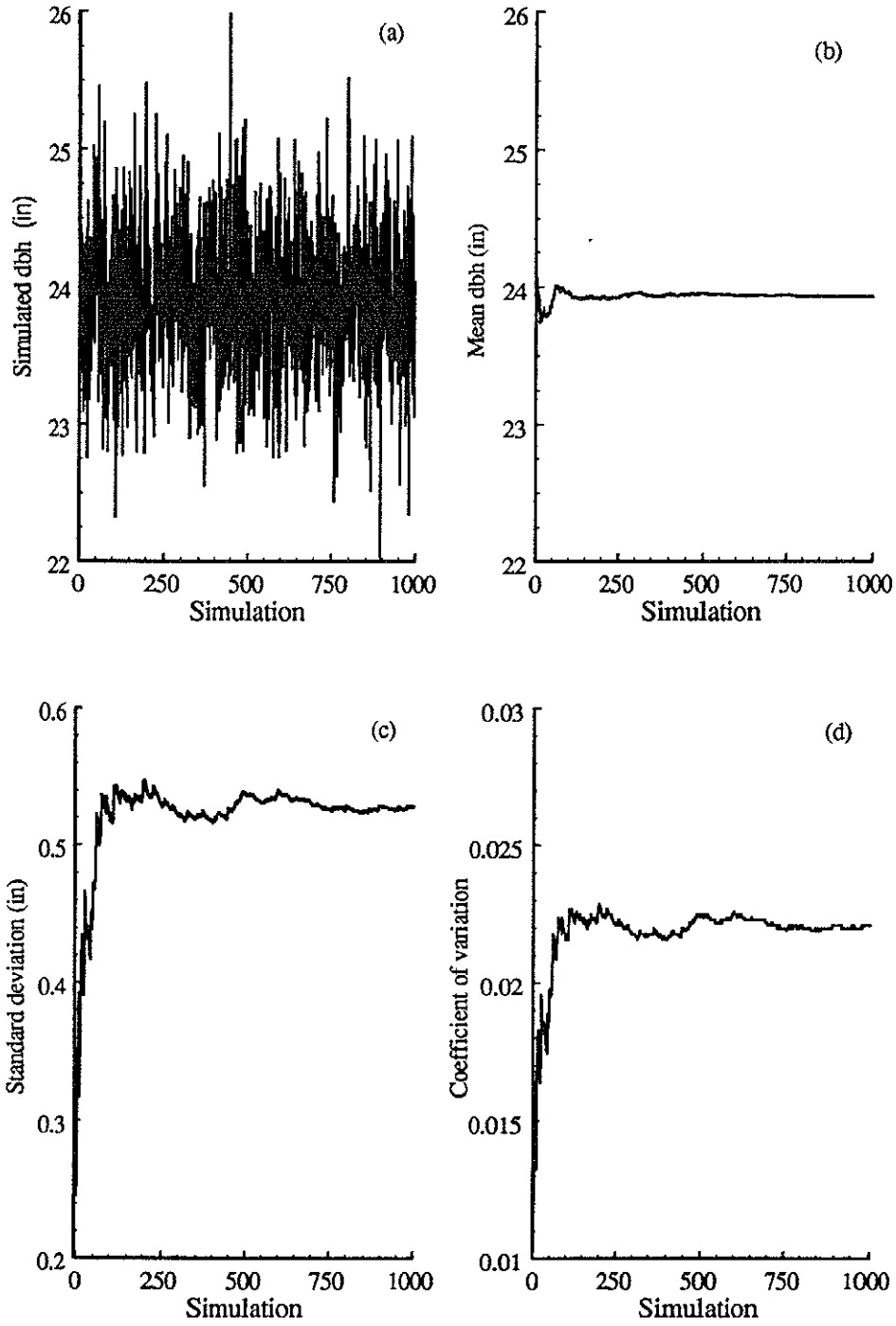

Figure 1.—*Simulation results for a sample tree: (a) simulated values of dbh after 20 years; (b) mean of simulated values of dbh after 20 years; (c) standard deviation of simulated values of dbh after 20 years; and (d) coefficient of variation of simulated values of dbh after 20 years.*

Table 1a.—*Median coefficients of variation for tree diameter estimates using STEMS-FIA*

Model calibration variables	Sources of uncertainty				Coefficient of variation[1]		
	Res[2]	DBH	CR	SI	5-yr	10-yr	20-yr
All[3]	x				0.096	0.107	0.121
All	x	x			0.140	0.146	0.160
All	x	x	x		0.156	0.171	0.195
All	x	x		x	0.140	0.149	0.163
All	x	x	x	x	0.156	0.166	0.196

[1] Coefficient of variation: CV= $\hat{\sigma}/\hat{\mu}$.
[2] Residual variability resulting from model calibration and applied using [1].
[3] DBH, CR, SI, AD, BA.

Table 1b.—*Median coefficients of variation for tree diameter estimates using AFIS*

Model calibration variables	Sources of uncertainty				Coefficient of variation[1]		
	Res[2]	DBH	CC	CR	5-yr	10-yr	20-yr
All[3]	x				0.094	0.117	0.138
All	x	x			0.161	0.179	0.198
All	x	x	x		0.168	0.185	0.204
All	x	x		x	0.225	0.235	0.294
All	x	x	x	x	0.214	0.258	0.314
All except CR	x	x	x		0.161	0.181	0.186
DBH only	x	x			0.139	0.143	0.138

[1] Coefficient of variation: CV= $\hat{\sigma}/\hat{\mu}$.
[2] Residual variability resulting from model calibration and applied using [1].
[3] DBH, CR, CC, BAL, BA, AD, PHY, NT.

Table 2a.—*Median coefficients of variation for plot basal area estimates using STEMS-FIA*[1]

Model calibration variables	Sources of uncertainty				Coefficient of variation		
	Res	DBH	CR	SI	5-yr	10-yr	20-yr
All	x				0.055	0.061	0.056
All	x	x			0.069	0.069	0.065
All	x	x	x		0.078	0.097	0.111
All	x	x		x	0.085	0.108	0.129
All	x	x	x	x	0.103	0.132	0.160

[1] See footnotes at the end of table 1a.

Table 2b.—*Median coefficients of variation for plot basal area estimates using AFIS*[1]

Model calibration variables	Sources of uncertainty				Coefficient of variation		
	Res	DBH	CC	CR	5-yr	10-yr	20-yr
All	x				0.051	0.061	0.064
All	x	x			0.071	0.075	0.079
All	x	x	x		0.061	0.075	0.074
All	x	x		x	0.085	0.120	0.147
All	x	x	x	x	0.082	0.112	0.155
All except CR	x	x	x		0.076	0.079	0.071
DBH only	x	x			0.080	0.088	0.096

[1] See footnotes at the end of table 1b.

Table 3a.—Coefficients of variation for regional basal area estimates using STEMS-FIA[1]

Model calibration variables	Sources of uncertainty				Coefficient of variation					
					Basal area growth			Basal area		
	Res	DBH	CR	SI	5-yr	10-yr	20-yr	5-yr	10-yr	20-yr
All	x				0.010	0.007	0.004	0.003	0.004	0.003
All	x	x			0.008	0.006	0.005	0.003	0.004	0.003
All	x	x	x		0.011	0.009	0.009	0.004	0.005	0.007
All	x	x		x	0.023	0.022	0.022	0.008	0.011	0.015
All	x	x	x	x	0.022	0.021	0.021	0.008	0.011	0.015

[1] See footnotes at the end of table 1a.

Table 3b.—*Coefficients of variation for regional basal area estimates using AFIS*[1]

Model calibration variables	Sources of uncertainty				Coefficient of variation					
					Basal area growth			Basal area		
	Res	DBH	CC	CR	5-yr	10-yr	20-yr	5-yr	10-yr	20-yr
All	x				0.010	0.006	0.004	0.003	0.003	0.003
All	x	x			0.010	0.008	0.005	0.004	0.005	0.004
All	x	x	x		0.008	0.005	0.004	0.004	0.004	0.003
All	x	x		x	0.013	0.013	0.013	0.005	0.007	0.010
All	x	x	x	x	0.014	0.013	0.014	0.005	0.007	0.010
All except CR	x	x	x		0.010	0.007	0.003	0.004	0.004	0.003
DBH only	x	x			0.005	0.004	0.003	0.004	0.003	0.003

[1] See footnotes at the end of table 1b.

sources of uncertainty. Third, the relative importance of a variable is evaluated by comparing the uncertainty of predictions for two conditions: (1) the variable is both included in the model calibration and included as a source of uncertainty in the simulations; and (2) the variable is excluded from the model calibration and, therefore, is excluded as a source of uncertainty in the simulations. If the uncertainty of predictions associated with (2) is less than with (1), then serious consideration ought to be given to excluding the variable from the model, even if its inclusion significantly increases the quality of fit of the model to the data. In these simulations, no consideration was given to correlations among the measurement errors or uncertainties for different sources.

Models

For more than a decade, the Forest Inventory and Analysis program at the North Central Research Station has used the STEMS diameter growth models (Belcher *et al.* 1982) to predict the current diameter of trees on plots not measured in the current year. Although the STEMS models are used in FIA applications, they were originally calibrated using data obtained from research plots. Holdaway (2000) documents some of the differences in these two sources of data. As a means of assessing the effects of applying the STEMS models to data that differ from those from which they were developed, we recalibrated the STEMS models for several Lake States species using FIA data (Lessard 2000) and designated the results as the STEMS-FIA models. Predictions with both STEMS and STEMS-FIA are obtained as products of two components, a potential growth component and a modifier component. The potential component, calibrated from observations of dominant and co-dominant trees only, uses current diameter at breast height (dbh), stand site index (SI), and crown ratio (CR) to predict the potential diameter growth of a tree. The modifier component, calibrated from trees of all crown classes, reduces growth from the potential using measures of competition based on stand basal area (BA), maximum observed basal area for the species (BAmax), and the ratio of dbh to average stand diameter (AD). Some of the STEMS predictor variables, notably crown class, crown ratio, and site index, have been demonstrated to have non-negligible uncertainty.

Advances in statistical estimation methodology since development of the STEMS models in the late 1970's and early 1980's have led to better modeling techniques and to greater understanding of the negative effects of uncertainty in predictor variables. Thus, as a part of our effort to develop annual forest inventory procedures, we are constructing new diameter growth models based on slightly different concepts and using more current statistical methodology (Holdaway 2000). These models are designated AFIS for their intended application in an annual forest inventory system. Predictions with the AFIS models are obtained as products of a regional median growth component expressed as a three-parameter Weibull-type function (Yang *et al.* 1978) of dbh and a modifier component expressed as the exponential of a linear combination of factors reflecting departures from regional median conditions for CR, BA, AD, crown class (CC), physiographic class (PHY), number of trees per acre (NT), and basal area in trees larger than the subject tree (BAL).

Data

The STEMS-FIA and AFIS models were both calibrated using FIA data obtained from the 1977 and 1990 periodic inventories in northeastern Minnesota for red pine (*Pinus resinosa* Ait.), balsam fir (*Abies balsamea* (L.) Mill.), quaking aspen (*Populus tremuloides Michx.*), and paper birch (*Betula papyrifera* Marsh.). The data used to initialize the simulations consist of plot- and tree-level observations obtained for FIA Unit 2 in northeastern Minnesota from the 1990 inventory. Because both the STEMS-FIA and AFIS models predict diameter growth for individual trees within the competitive environment of all trees growing on a plot, the predicted growth for an individual tree depends to some degree on the predicted growth of all trees on the plot. Thus, the data used to initialize the simulations were restricted to observations for plots consisting of only the four species for which the models had been calibrated. This data set consisted of observations for 441 trees growing on 20 plots.

RESULTS

Tree Diameter Estimates

When the models were calibrated using the full set of predictor variables, the uncertainty in initial plot and tree conditions substantially increased the uncertainty of model diameter predictions for both the STEMS-FIA and AFIS models (table 1a-b) (fig. 2). For example, the median coefficient of variation for 20-year AFIS predictions increased from 0.138 when only residual variability was considered to 0.314 when uncertainty from all sources was considered. At all time intervals, median values of coefficients of variation were less for the STEMS-FIA models than for the AFIS models when both models were calibrated with the full set of predictor variables and when all sources of uncertainty were used to estimate model prediction uncertainty. Although residual variability and dbh measurement error had the greatest individual effects on the uncertainty of model predictions for both models, neither of these sources can be eliminated; residual variability is unavoidable, and dbh is fundamental to the underlying model. Of the remaining sources of uncertainty, crown ratio had the single greatest detrimental effect for both models. As a test of the overall

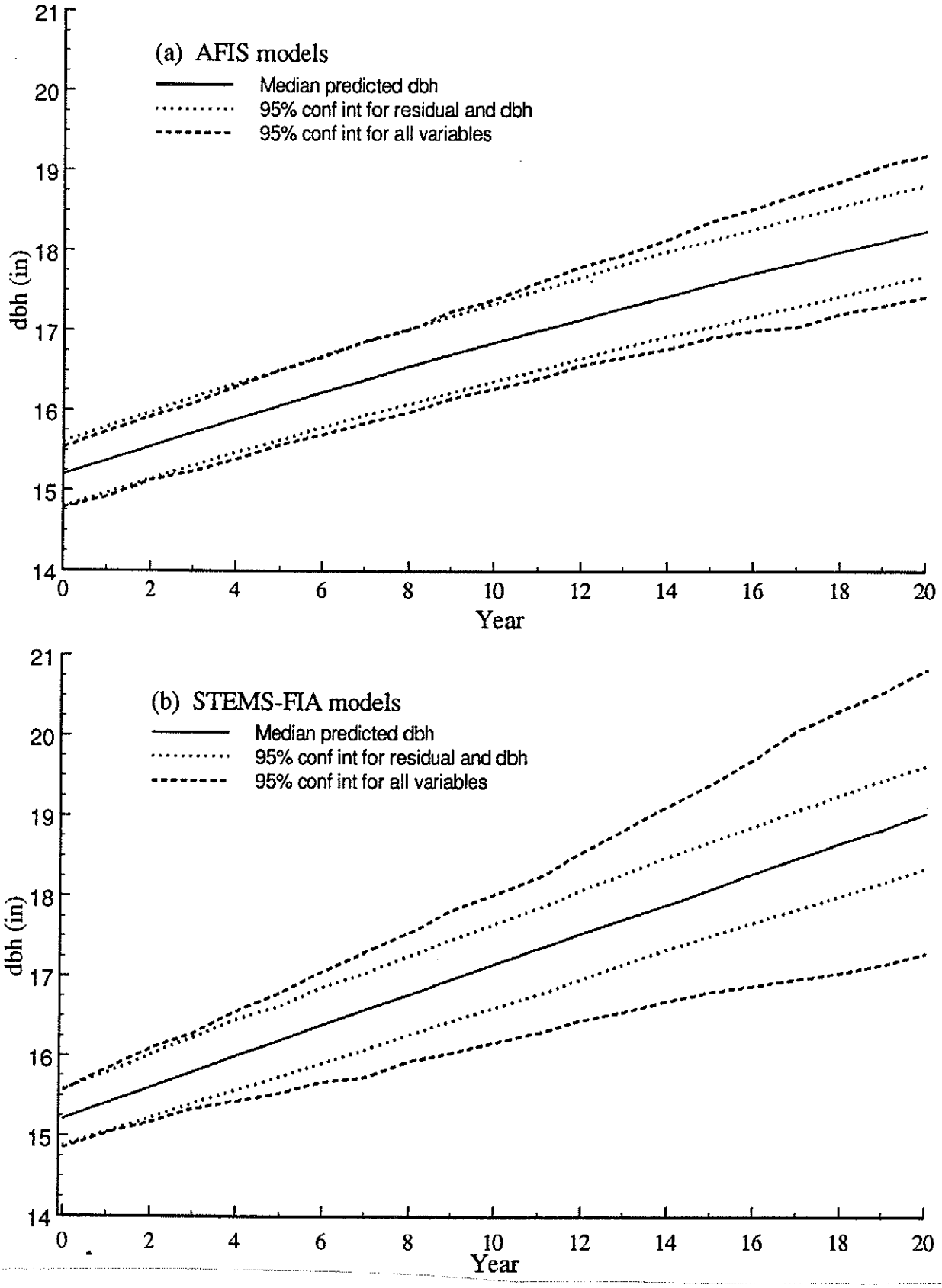

Figure 2.—*Ninety-five percent confidence intervals for predicted dbh of a sample tree; models calibrated with all variables (see footnotes on table 1a-b); confidence interval based on uncertainty from indicated sources: (a) AFIS; (b) STEMS-FIA.*

effect of this variable, the AFIS models were calibrated without using crown ratio as a predictor, a new relationship between predicted growth and residual variability as per [1] was estimated, and the simulations were repeated. The result of eliminating crown ratio as a predictor variable and as a source of uncertainty in initial conditions was a decrease in the uncertainty of predictions (table 1b). Without consideration of bias, this result suggests that although crown ratio was a significant predictor variable when fitting the model to the calibration data, the overall effect of the variable was an increase in the model prediction uncertainty. Similarly, model prediction uncertainty was reduced even more when all predictor variables except dbh were deleted from the model (table 1b).

Plot-level Basal Area Per Acre Estimates

At the plot level, the uncertainty in predictor variables also had a negative effect on the uncertainty of predictions for both models when they were calibrated with the full set of predictor variables (table 2a-b). However, in contrast to the tree-level results, uncertainty in plot-level estimates was somewhat less for the AFIS models. This result is attributed to the negative effects of uncertainty in site index, a plot-level predictor variable used in the STEMS-FIA models but not in the AFIS models. As with the diameter predictions, prediction uncertainty for the AFIS models was less when dbh was the only predictor variable than when all predictor variables were included in the model.

Regional Basal Area and Basal Area Growth Estimates

Uncertainty in predictor variables had a negative effect on the uncertainty of regional estimates of both basal area and basal area growth, but the effect was not nearly as severe as for tree- and plot-level estimates (table 3a-b). At the regional level, the AFIS models produced estimates with less uncertainty than did the STEMS-FIA models, a result that is again attributed to the effects of uncertainty in the plot-level predictor variable, site index, in the STEMS-FIA models. As with the tree- and plot-level results, removal of CR from the set of predictor variables reduced the uncertainty of AFIS predictions; the greatest reduction was achieved when all predictor variables except dbh were removed from the model.

CONCLUSIONS

Six immediate conclusions appear evident: (1) uncertainty in initial plot- and tree-level conditions has a substantial, negative effect on the uncertainty of model predictions; (2) serious consideration and additional analyses are needed to determine if predictor variables with large uncertainties should be included in models; (3) although the uncertainty in diameter growth predictions for individual trees was relatively large, the effects of propagating this uncertainty to plot-level estimates of basal area per acre growth and to regional estimates of basal area were not severe; (4) without considering bias effects, eliminating all but the few most crucial predictor variables often produced the best results at the regional level; (5) the uncertainty of STEMS-FIA predictions at plot- and region-level were more sensitive to uncertainty in the plot-level predictor variable site index than to uncertainty in tree-level predictor variables; and (6) the AFIS models were as good or better than the STEMS-FIA models with respect to the uncertainty of plot and regional predictions.

Several broad conclusions also emerge. First, even when the only the source of uncertainty considered is residual variability, coefficients of variation were on the order of 0.10 or greater for tree diameter predictions, suggesting that it may be extremely difficult to develop precise individual tree diameter growth models when the calibration data are obtained from trees representing a broad geographic region. In addition, if uncertainty in parameter estimates had been included in the analyses, these coefficients of variation would have been even larger. Second, despite the uncertainty in diameter predictions, coefficients of variation on the order of 0.01 or less for regional basal area estimates suggest that reasonably precise estimates may be obtained at this level. Although these results will deteriorate when parameter covariances are included, this result is still encouraging. Finally, the conclusions suggest that for broad regional applications, predictions from simpler models with fewer predictor variables that capture basic growth relationships may be less uncertain than those from more complex models with many predictor variables that attempt to explain a large proportion of variation.

ACKNOWLEDGMENTS

The following people reviewed this manuscript: Gary Brand and Mark Hansen, USDA Forest Service, North Central Research Station, St. Paul, MN, USA, and H. Todd Mowrer, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA.

LITERATURE CITED

Belcher, D.W.; Holdaway, M.R.; Brand, G.J. 1982. A description of 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. 18 p.

Fuller, W.A. 1987. Measurement error models. New York, NY: John Wiley and Sons. 440 p.

Holdaway, M.R. 2000. A FIA tree growth model for updating annual forest inventories in Minnesota. In: Hansen, M.; Burk, T., eds. Integrating tools for natural resources inventories in the 21st century; 1998 August 16-20; Boise, ID. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 507-514.

Lessard, V.C. 2000. Calibration of the STEMS diameter growth model on FIA data. In: Hansen, M.; Burk, T., eds. Integrating tools for natural resources inventories in the 21st century; 1998 August 16-20; Boise, ID. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 525-532.

McRoberts, R.E.; Hahn, J.T.; Hefty, G.J.; Van Cleve, J.R. 1994. Variation in forest inventory field measurements. Canadian Journal of Forest Research. 24: 1766-1770.

McRoberts, R.E. 1996. Estimating variation in field crew estimates of site index. Canadian Journal of Forest Research. 26: 560-565.

Nichols, N.S.; Gregoire, T.G.; Zedaker, S.M. 1991. The reliability of tree crown position classification. Canadian Journal of Forest Research. 21: 698-701.

Yang, R.C.; Kozak, A.; Smith, J.H.G. 1978. The potential of Weibull-type functions as flexible growth curves. Canadian Journal of Forest Research. 8: 423-431.