ANALYZING THE UNCERTAINTY OF DIAMETER GROWTH MODEL PREDICTIONS

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

McRoberts, Ronald E.; Lessard, Veronica C.; Holdaway, Margaret R. 1999. Analyzing the uncertainty of diameter growth model predictions. In: Proceedings, Society of American Foresters 1998 national convention; 1998 September 19-23; Traverse City, MI. SAF Publ. SAF 99-01. Bethesda, MD: Society of American Foresters: 319-324.

ABSTRACT. The North Central Research Station of the USDA Forest Service is developing a new set of individual tree, diameter growth models to be used as a component of an annual forest inventory system. The criterion for selection of predictor variables for these models is the uncertainty in 5-, 10-, and 20-year diameter growth predictions estimated using Monte Carlo simulations. Particular attention is focused on the uncertainty in the model predictions that can be attributed to five sources of uncertainty: (1) residual variation after model calibration; (2) covariances for parameter estimates; (3) measurement error for diameter at breast height; (4) measurement error for crown ratio; and (5) measurement error for crown class. The cumulative effects of uncertainty in model predictions for individual trees on the uncertainty of basal area estimates at the plot and regional levels are also investigated. For four Lakes States species, the results suggest that uncertainty in predictor variables has a non-negligible effect on the uncertainty of diameter predictions, but that these effects are attenuated when results are aggregated at the regional level.
KEY WORDS. Monte Carlo simulations, annual forest inventory

INTRODUCTION

The North Central Research Station (NCRS) currently uses the STEMS (Belcher et al, 1982) diameter growth models in its periodic inventories to update the status of trees on Forest Inventory and Analysis (FIA) plots not measured in the current cycle. In addition, NCRS is developing new diameter growth models for use in an annual forest inventory system. Usually the criterion for selection of predictor variables for such models is 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 its values, as is frequently the case with ocular measurements, or when its values are estimates based on samples. When such predictor variables are included in models, the effects of their uncertainties may substantially increase the total uncertainty of model predictions. Thus, the objectives of the study are to estimate the total uncertainty in diameter growth model predictions and to investigate the effects of including model predictor variables that have non-negligible uncertainty.

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 time intervals, and repeat the process a large number of times. Plot-level estimates of basal area per acre are obtained by summing results over all trees on plots, and regional estimates of basal area are obtained by summing results over all plots. Because the regional basal area estimates are based on results from only 20 plots (see METHODS-Data section), they do not accurately represent true regional basal area. However, these estimates permit investigation of uncertainty issues 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 for it in this study.

A Monte Carlo process is used to simulate uncertainty from 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 values of predictor variables used in calibrating the model and uncertainty in the initial conditions to which the models are applied are negligible, then the preceding sources of variation are the only two that need to be considered when estimating model prediction uncertainty. However, when uncertainty in values of the predictor variables and in initial tree conditions is non-negligible, then this uncertainty must be incorporated into the model parameter covariances and propagated through the models. Thus, the Monte Carlo process is also used to simulate uncertainty for three additional sources: (3) crown class observation error; (4) crown ratio observation error; and (5) diameter measurement error. Note that uncertainty in diameter measurements generates additional uncertainty in variables derived from diameter such as stand basal area. In addition, when sampling is conducted for variable radius plots, diameter measurement error affects tree expansion factors and, hence, estimates of numbers of trees per acre.

Distributions representing the uncertainty from the above five sources were obtained when calibrating the models or from the literature. Estimates of residual variation were obtained as by-products of calibrating the models (Holdaway 1998). Residual variation is assumed to be normally distributed but heterogeneous with standard deviations related to predicted growth as follows:

$$E[\ln(\sigma_\epsilon)]=\beta_1+\beta_2\ln(\Delta d), \quad [1]$$

where E(.) represents statistical expectation, Δd is annual diameter growth prediction, σ_ϵ is the standard deviation of the residuals for predicted diameter growth classes of small widths, and the βs are parameters to be estimated. Because the parameter covariance estimates for nonlinear models obtained using analytical methods are known to be unreliable, the uncertainty in parameter estimates for this study was obtained using a separate five-step Monte Carlo process: (1) simulated values of the predictor variables were obtained by adding their observed values and randomly generated uncertainty based on the measurement error distributions obtained from the literature; (2) the parameter estimates obtained from calibrating the models to the original data were used to calculate predictions using the simulated values of the predictor variables obtained in step 1; (3) residual variation was added to these predictions to obtain simulated observations; (4) new parameter estimates were obtained by calibrating the models to the simulated data generated in steps 1-3; and (5) steps 1-4 were repeated a large number of times to generate a distribution of parameter estimates. Nichols et al (1991) reported that the repeatability of crown class observations by field crews in the same year is approximately 80% with the other 20% allocated uniformly to the two adjacent classes. Finally, 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, and they provided a mathematical model for estimating diameter measurement error as a function of diameter.

The Monte Carlo process for estimating model prediction uncertainty involves six steps: (1) mimic uncertainty in initial tree and plot conditions by adding randomly generated variation to tree- and plot-level variables such as crown ratio, crown class, and diameter; (2) randomly select a set of parameter estimates from the distribution previously generated; (3) use the models to predict annual diameter growth for each tree, mimic residual uncertainty in the predictions with randomly generated variation, and obtain simulated 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 estimates for plots, and basal area estimates for the region at the 5-, 10-, and 20-year intervals; and (6) replicate steps 1-5 a large number of times. No consideration is given to correlations among the measurement errors or uncertainties for different sources. In step 5, the basal area per acre estimates are obtained by multiplying the tree basal area estimates and the corresponding tree expansion factors for all trees and then summing these products for all trees on each plot. Relative estimates of basal area at the regional level are obtained by adding basal area per acre 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.

Models

The FIA program at the NCRS currently uses the STEMS diameter growth models (Belcher et al 1982) for its periodic inventory system to predict the current diameter of trees not measured in the current cycle. Although the STEMS models generally have been accepted as sufficiently accurate at the regional level, several of their features bear further consideration: (1) they were calibrated using data from research plots but are applied to FIA data (Holdaway 1998); (2) accurate values for some of their predictor variables are known to be difficult to obtain; (3) they are based on the concept of potential growth which cannot be observed; (4) improvements in statistical estimation methodology have been developed since they were calibrated; and (5) the uncertainty of their model predictions cannot be readily obtained. Thus, we are constructing new diameter growth models based on slightly different concepts and using more current statistical methodology (Holdaway 1998). These models are designated AFIS for their intended application in an annual forest inventory system. Diameter growth 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 diameter at breast height (DBH) and a modifier component expressed as the exponential of a linear combination of factors reflecting departures from regional median conditions selected from among crown ratio (CR), basal area per acre (BA), basal area per acre in trees larger than the subject tree (BAL), crown class (CC), physiographic class (PHY), average stand diameter (AD), site index (SI), and number of trees per acre (NT). Comparisons of the STEMS and AFIS model with respect to several factors have been previously reported (Holdaway 1998, Lessard 1998, McRoberts *et al* 1998).

Data

The AFIS models were 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 the 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 on the predicted growth of all trees on the plot. Thus, the data used to initialize the simulations was 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.

Analyses

Distributions of the coefficients of variation ($CV=\hat{\sigma} / \hat{\mu}$) obtained from 5000 replications of the simulation process serve as the basis for analyzing the uncertainty of model predictions. Medians, means, standard deviations, and coefficients of variation for simulated DBH for individual trees stabilized by completion of 5000 simulations. The results of the simulations are summarized as median coefficients of variation for diameter estimates of individual trees, median coefficients of variation for plot-level estimates of basal area per acre, and coefficients of variation for regional estimates of basal area at 5-, 10-, and 20-year intervals (Tables 1). Summaries are expressed in terms of medians because of possible asymmetry in the distributions.

The uncertainty of model predictions is analyzed with respect to two factors. First, 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 sources of uncertainty. Second, 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 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.

RESULTS

Tree diameter estimates

For individual trees, the total uncertainty in model predictions ranged from 0.082 to 0.096 for 5-year predictions; from 0.085 to 0.115 for 10-year predictions, and from 0.081 to 0.143 for 20-year predictions. The increase in uncertainty as the prediction interval increases is as expected. The general ranges of these values translate to confidence interval widths ranging from approximately 15% to slightly less than 30% of the prediction (Table 1). The uncertainty of model diameter predictions increased substantially when the uncertainty in the initial tree conditions to which the models were applied was considered. For example, when uncertainty in the initial conditions were included in the analysis of uncertainty for 5-year model predictions, the median coefficients of variation for individual trees increased from 0.062 to 0.082 when DBH was the only predictor variable, from 0.061 to 0.094 when DBH, CR, and BAL were used as predictor variables, and from 0.065 to 0.095 when the largest number of predictor variables was included in the model. Overall, the unavoidable uncertainty due to residual variability and covariances of the parameter estimates contributed most to the uncertainty in the model predictions. Note that in all analyses, the effects of uncertainty in DBH, BA and BAL cannot be separated because they all arise due to uncertainty in DBH measurements. Uncertainty in DBH measurements was the second greatest contributor to uncertainty in 5-year model predictions, while CR was the second greatest contributor for 10-year and 20-year model predictions. Although the models with DBH alone and with DBH and BAL as predictor variables had the smallest median coefficients of variation, no inference should be made that these are the most accurate models. First, these models account for considerably less variation in the observations than do models with more predictors, and second, these analyses do not include an assessment of bias. However, the finding that models with greater numbers of significant predictor variables sometimes also have greater prediction uncertainty suggests that quality of fit of a model to data may not be the best criterion for selection of predictor variables when those variables have uncertainty associated with them.

Plot-level basal area per acre estimates

At the plot level, median coefficients of variation for estimates of basal area per acre across all prediction intervals ranged from 0.016 to 0.028, considerably less than for diameter predictions for individual trees (Table 1). The uncertainty in initial conditions had a negative effect on the uncertainty of basal area per acre estimates, but the effect was not as great as for individual tree diameter predictions. The single greatest contributor to the uncertainty in basal area per acre estimates, as for diameter predictions for individual trees, was the joint uncertainty due to residual variability and parameter covariances, neither of which can be eliminated. The second greatest contributor was CR, although for 5-year predictions the effects of uncertainty in CR and DBH were similar.

Regional basal area estimates

For regional basal area estimates, the largest coefficients of variation were only slightly greater than 0.01 for 5-year estimates and only slightly greater than 0.02 for 20-year estimates (Table 1). Although the effects of uncertainty in initial conditions were again detrimental, the magnitude of the effects was much less than for either tree diameter predictions or plot basal area per acre estimates. In fact, the magnitudes were so small that an ordering of the individual sources of uncertainty with respect to their effects on the uncertainty of model predictions is not feasible.

CONCLUSIONS

Four immediate conclusions appear evident: (1) uncertainty in initial tree 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 basal area per acre estimates and to regional basal area estimates were not severe; and (4) without considering bias effects, eliminating all but the few most crucial predictor variables often produced better results.

Several broad conclusions also emerge. First, even when the only the source of uncertainty considered is residual variability and parameter covariances, coefficients of variation were on the order of 0.10±0.04 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 large geographic region. 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. 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 large numbers of predictor variables that attempt to explain a large proportion of variation.

LITERATURE CITED

BELCHER, D.W., HOLDAWAY, M.R., and BRAND, G.J. 1982. A description of STEMS - the stand and tree evaluation and modeling system. USDA For. Serv. Gen Tech. Rep. North Cent. For. Exp. Stn., St. Paul, MN 55108.

HOLDAWAY, M.R. 1998. A FIA tree growth model for updating annual forest inventories in Minnesota. In *Integrating tools for natural resources inventories in the 21st century*. Ed.: Hansen, M., Fairweather, S., and Burk, T. Proceedings of the IUFRO conference, August 16-20, 1998. Boise, Idaho, USA. USDA For. Serv. Gen Tech. Rep. North Cent. Res. Stn., St. Paul, MN 55108, USA (in press).

LESSARD, V.C. 1998. Calibration of the STEMS diameter growth model on FIA data. In *Integrating tools for natural resources inventories in the 21st century*. Ed.: Hansen, M., Fairweather, S., and Burk, T. Proceedings of the IUFRO conference, August 16-20, 1998. Boise, Idaho, USA. USDA For. Serv. Gen Tech. Rep. North Cent. Res. Stn., St. Paul, MN 55108, USA (in press).

McROBERTS, R.E., HAHN, J.T., HEFTY, G.J. and VAN CLEVE. 1994. Variation in forest inventory field measurements. *Canadian Journal of Forest Research* 24:1766-1770.

McROBERTS, R.E., HOLDAWAY, M.R., and LESSARD, R.C. 1998. Comparing the STEMS and AFIS growth models with respect to the uncertainty of predictions. In *Integrating tools for natural resources inventories in the 21st century*. Ed.: Hansen, M., Fairweather, S., and Burk, T. Proceedings of the IUFRO conference, August 16-20, 1998. Boise, Idaho, USA. USDA For. Serv. Gen Tech. Rep. North Cent. Res. Stn., St. Paul, MN 55108, USA (in press).

NICHOLS, N.S., GREGOIRE, and 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., and SMITH, J.H.G. 1978. The potential of Weibull-type functions as flexible growth curves. *Canadian Journal of Forest Research* 8:423-431.

ABOUT THE AUTHORS

The authors are mathematical statistician, research forester, and mathematical statistician, respectively, at the North Central Research Station, USDA Forest Service, St. Paul, MN 55108, USA.