American Statistical Association

1999
Proceedings

of the

Section on Statistics and the Environment

Papers presented at the
Annual Meeting of the American Statistical Association,
Baltimore, Maryland, August 8–12, 1999,
and the
Ninth Lukacs Symposium,
Bowling Green State University
Bowling Green, Ohio
April 23–25, 1999
under the sponsorship of the
Section on Statistics and the Environment

American Statistical Association ■ 1429 Duke Street, Alexandria, VA 22314

ESTIMATING UNCERTAINTY IN ANNUAL FOREST INVENTORY ESTIMATES

Ronald E. McRoberts, Veronica C. Lessard, USDA Forest Service
Ronald E. McRoberts, North Central Research Station, 1992 Folwell, Avenue, St. Paul, MN 55108

Key Words: Monte Carlo simulation, Forest inventory, Growth model, Moving average

ABSTRACT: The precision of annual forest inventory estimates may be negatively affected by uncertainty from a variety of sources including: (1) sampling error; (2) procedures for updating plots not measured in the current year; and (3) measurement errors. The impact of these sources of uncertainty on final inventory estimates is investigated using Monte Carlo simulation techniques.

INTRODUCTION

Forest Inventory and Analysis (FIA) is a continuing endeavor of the USDA Forest Service as mandated by the Renewable Resources Act of 1978 and previously by the McSweeney-McNary Forest Research Act of 1928. The objective of FIA has been to periodically inventory the forest land of the United States to determine its extent and condition and the volume of standing timber, timber growth, and timber depletions. USDA Forest Service regional research stations are responsible for conducting these inventories and publishing summary reports for individual states. Periodic, statewide inventories have been conducted since the 1930s at recent intervals of 11-17 years. The precision of periodic inventory estimates is degraded by the effects of conducting inventories over multiple years, and it further decreases over time due to factors such as changes in land use and tree growth, mortality, and removals between inventories.

With passage of the 1998 Farm Bill, formally known as the Agricultural Research Extension and Education Reform Act of 1998, the US Congress has required the USDA Forest Service to conduct annual forest inventories throughout the nation with 20% of plots to be measured every year and resource reports to be published every 5 years. However, the 20% annual sample requirement may be relaxed, particularly in the western states.

ANNUAL SAMPLE

Forest inventories are described as annual on the basis of the measurement of an annual sample of plots and the capability of annually producing population estimates; the basis is not a complete annual inventory of all plots. FIA precision standards require that standard errors of estimates not exceed 5% per billion cubic feet of growing stock or 3% per million acres of forest land area (USDA, Forest Service 1970). These standards have been found to require a sampling intensity of one plot for approximately every 6,000 acres. To satisfy this requirement, a grid of non-overlapping hexagons, each approximately 5,900 acres in area, was constructed by dividing the similar, but larger, hexagons established for the Forest Health Monitoring (FHM) program (Eagar *et al* 1991, White *et al* 1992).

In each 5,900-acre hexagon, the existing FIA plot nearest the center was selected, or if no FIA plot existed in the hexagon, a new plot was established at the hexagon center. This network of plots is designated as the federal base sample and its measurement is funded by USDA Forest Service. The federal base sample is considered an equal probability sample of the total surface area of a state with the basis for inference residing in the sample design. This assumption of equal probability sampling is based on the random orientation of the system of the FHM hexagons and the lack of relationship between the locations of the hexagons and the locations of permanent FIA field plots. Thus, when the distribution of the variables in the population is essentially random, simply random sampling theory can be used in conjunction with a systematic sample design.

INVENTORY ESTIMATION

The federal base sample was systematically divided into five non-overlapping, interpenetrating panels to accommodate the 1998 Farm Bill requirement for a 20% annual sample. Each year the plots in a single panel are selected for measurement with panels selected on a 5-year, rotating basis. For estimation purposes, the measurement of each panel of plots is considered an independent sample of all lands in a state, and the cumulative measurement of all plots over the 5-year cycle is considered a single equal probability sample.

Annual forest inventories feature measurement of one panel of plots each year. Therefore, the simplest approach to calculating annual FIA estimates is to use only the data from the panel of plots measured in the current year. While these estimates reflect current conditions and are based entirely on measured plots, their precision may be unacceptable for some variables due to the small annual sample size. An alternative is to base the annual estimates on the most recent information for all plots. The advantage of this alternative is that precision is increased because all plots are used for estimation; the disadvantage is that the estimates do not reflect current conditions but rather a moving average of conditions over the past 5 years. In particular, the moving average will be biased for monotonic trends, will lag behind emerging trends, but will dampen the effects of temporary or sporadic changes. Another alternative is to first update to the current year information for all plots measured in previous years and then base the estimates on all plots. If the updates are unbiased and sufficiently precise, this alternative increases the precision of the estimates without the adverse effects of using out-of-date information.

MODEL UPDATING

For more than a decade the North Central Research Station (NCRS) of the USDA Forest Service has used diameter growth, mortality, and harvest removals models to update plots not measured in the current inventory cycle. Recent diameter growth modeling research at NCRS has focused on development of new individual tree, diameter growth models that address four factors: (1) calibration on FIA plot data; (2) negligible bias; (3) predictions of known precision; and (4) avoidance, where possible, of predictor variables that are sample estimates or are difficult to precisely measure.

Data. The data used to calibrate and validate the DBH growth models were taken from measurements of forested Minnesota FIA plots for the 1977 and 1990 inventories (Holdaway 1999). Only those trees alive and measured in both inventories were used. For each tree, average annual growth in diameter at breast height (DBH) is calculated as the ratio of the difference in DBH measurements for the two inventories and the number of years between measurements. Due to nonlinearity in the relationship between DBH and DBH growth, expected annual DBH growth between inventories may not be considered constant and is therefore imprecisely represented by average annual DBH growth between inventories (McRoberts 1998). However, because preliminary investigations indicated that the effects of the nonlinearity were small for this application, average annual DBH growth over the measurement interval was used as a surrogate for annual DBH growth.

Predictors of annual DBH growth were selected from a variety of tree and plot variables that included DBH, crown ratio (CR), crown class (CC), average plot diameter (AD), plot basal area (BA), plot basal area in trees larger than the subject tree (BAL), site index (SI), number of trees per acre (NT), plot physiographic class (PC), and relative diameter calculated as the ratio of DBH and AD (REL). Of these potential predictor variables, only DBH, CR, and CC correspond to the measurement of a single physical entity; AD, SI, and PC are sample-based estimates of site attributes with PC being related to soil and moisture conditions; and BA, BAL, and NT are sample-based per acre estimates. Although DBH may be quite precisely measured, both CR and CC are obtained by ocular estimation and are known to be subject to imprecision (Nicholas *et al* 1991, McRoberts *et al* 1994). Values of the remaining variables are all sample estimates and are therefore subject to sampling variability. Measurement error and sampling variability in predictor variables propagate through models and have an adverse effect on the precision of predictions. Thus, where possible, predictor variables with large uncertainties were avoided in model development, and where impossible, the effects of uncertainty in these variables on the precision of model predictions were quantified.

Model form. The DBH growth model consists of the product of two components, an average component corresponding to regional average DBH growth and a modifier component that adjusts DBH growth predictions in accordance with local plot or tree conditions. The average component is based on a 2-parameter gamma model with a constant multiplier and uses DBH as the predictor variable, while the modifier component consists of the product of exponentials, each of which incorporates a single predictor variable. Several forms were considered for the average component including the Chapman-Richards model, the Weibull model, and the gamma model. The gamma model was selected on the basis of superior quality of fit and more consistent convergence with the nonlinear model parameter estimation routine. Each factor in the modifier product expresses a multiplicative effect on growth in terms of departures from regional or ecosystem means for a single predictor variable. Thus, the form of the DBH growth model is

$$E(\Delta DBH) = AVE(DBH)*MOD(X_1, \ldots X_5) \qquad [1a]$$

where the βs are parameters to be estimated, E(.) is statistical expectation, ΔDBH is annual DBH growth;

$$AVE(DBH) = \beta_1 \exp(-\beta_2\, DBH)\, DBH^{\beta_3}, \qquad [1b]$$

and

$$MOD(X_1,\ldots,X_5)=\prod_{i=1}^{5} e^{f_i(X_i)}, \qquad [1c]$$

with the X_is and the f_is as identified in Table 1.

Table 1. Predictor variables and functions for [1c].

i	X_i	$f_i(x_i)$
1	CR	β_4 (CR-4)
2	BAL	β_5 (BAL-50)
3	BA	β_6(BA-100)
4	CC	$\beta_7 (e^{CC/3} - 2.178)$
5	PC	$\beta_8(PC-5)+\beta_9(PC-5)^2$

Parameter estimation. Model [1] was fit to the data for each species separately. Predictor variables in the MOD component of the model, as expressed by [1c] and Table 1, were selected for inclusion in the model using a comparison criterion described by Linhart and Zucchini (1986) that assesses improvement in quality of fit in terms of both changes in residual sums of squares and number of model parameters. If a predictor variable was not selected for inclusion in the model, the corresponding parameter estimate in Table 1 is simply set to zero.

Due to heterogeneity of variance, iterative reweighted nonlinear regressions were used to obtain parameter estimates. The relationship between residual standard deviations, σ_{res}, and predictions of annual diameter growth, ΔDBH, was found to be adequately described by

$$E[\ \ln(\sigma_{res})\ ]=\alpha_1+\alpha_2 \ln(\ \Delta DBH\ ), \qquad [2]$$

where E(.) is statistical expectation, ln(.) is the natural logarithm function, and the αs are coefficients estimated from the data. Thus, weights for each subsequent weighted nonlinear regression were obtained by inverting the squared predictions of σ_{res} obtained from [2] for the previous regression.

COMPARING ESTIMATION ALTERNATIVES

As previously noted, three estimation alternatives are available for the analyses of annual forest inventory data: (1) calculate inventory estimates using data from the most current panel only; (2) calculate a moving average over the past 5 years of panel data; and (3) calculate annual estimates using the current panel of data and past panels of data that have been updated to the current year using growth, mortality, and harvest removals models. The bias and precision of basal area per acre estimates using these three alternatives were compared using a Monte Carlo simulation procedure based on a subset of data for FIA plots in Minnesota Inventory Units 1 and 2. The subset consisted of data for the 313 plots whose species composition included only the fifteen most frequently occurring species in Minnesota.

Precision of model predictions. To accurately estimate precision of basal area per acre estimates using the model updating alternative, the precision of the model growth predictions must be first estimated. Due to the extreme complexity of this task using analytical methods, Monte Carlo simulations processes were used to simulate uncertainty from three sources: residual variability around an estimated regression line; covariances of model parameter estimates; and uncertainty in the form of measurement error and sampling variability in the values of predictor variables. Distributions of the residuals were obtained as by-products of calibrating the models and are assumed to be $N(0,\sigma_{res})$ where σ_{res} is obtained from [2].

For this application, uncertainty in the values of the predictor variables is manifested twice: first, it increases model parameter covariance estimates because of uncertainty in predictor variables in the calibration data set, and second, it increases uncertainty in model predictions because of the uncertainty in the conditions of the trees and plots to which the model is applied. Thus, uncertainty in the predictor variables must be addressed before estimates of the model covariances can be obtained. Distributions representing the uncertainty in DBH, CR, and CC were obtained from the literature. McRoberts *et al* (1994) provided a mathematical model for estimating DBH measurement error as a function of diameter. Note that uncertainty in DBH generates additional uncertainty in variables derived from DBH such as BA and BAL. In addition, when sampling is conducted for variable radius plots, DBH measurement error affects tree expansion factors. McRoberts *et al* (1994) also reported that observations of CR for the same tree by separate field crews ranged ± 0.3 from the median. Finally, Nicholas *et al* (1991) reported that the repeatability of CC observations by field crews in the same-year is approximately 80% with the other 20% allocated uniformly to the two adjacent classes. Uncertainty for PC was not included in the simulations due to lack of information on sampling variability for this predictor variable.

Because parameter covariance estimates for nonlinear models obtained using analytical methods are known to be unreliable, distributions of parameter estimates that more accurately represent the covariance relationships were generated using a 5-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 and sampling variability distributions obtained from the literature;

2. 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. Randomly generated residual variation, with standard deviation calculated using [2], 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;

5. Steps 1-4 were repeated a large number of times to generate a distribution of parameter estimates.

Precision of basal area per acre estimates. A Monte Carlo process for estimating the uncertainty in model predictions and the effects of that uncertainty on inventory estimates was designed to mimic the NCRS annual inventory procedure. The simulation procedure featured measurement of a rotating 20% sample of plots each year, annual updating of the non-measured plots using the models, and calculation of mean basal area per acre each year using each of the three estimation alternatives. The simulation procedure included no effects of regeneration, mortality, harvest, other removals or conversion of forest land to other cover types.

The simulation procedure was in two phases, with Phase 1 simulating observed annual data and Phase 2 simulating the FIA sampling, model updating, and estimation processes:

Phase 1:

1. For Year 0, mimic uncertainty in initial tree and plot conditions by adding randomly generated measurement error to the observed initial tree conditions expressed by the predictor variables, DBH, CR, CC, BA, and BAL; record the year and the mean and variance of basal area per acre across all plots;

2. Use the models with the parameter estimates obtained from the original calibration data set to predict annual DBH growth for each tree; mimic residual uncertainty in the predictions with randomly generated variation; add DBH at the beginning of the year, predicted annual DBH growth, and residual variation to obtain DBH for the beginning of the subsequent year;

3. Update plot-level variables BA and BAL using new DBH values; record the year and mean and variance of basal area per acre across all plots;

4. Repeat Steps 2-3 for 20 years.

Phase 2:

1. Randomly select without replacement a set of parameter estimates for the models from among those generated with the previously described 5-step Monte Carlo process.

2. For Year 0, mimic uncertainty in initial tree and plot conditions by adding randomly generated measurement error to the observed initial tree conditions expressed by the predictor variables, DBH, CR, CC, BA, and BAL; calculate the mean and variance of basal area per acre across all plots;

3. For years subsequent to Year 0, and for the 20% of plots to be measured in the current year, replace the previous year's DBH for each tree with the Phase 1 values; add randomly generated uncertainty to mimic measurement error; update BA and BAL for these plots for the current year; record the year, tree DBHs, and plot basal area per acres estimates;

4. For years subsequent to Year 0 and for the 80% of plots not measured in the current year, use the models with parameter estimates selected in Phase 2, Step 1 to predict annual DBH growth for each tree; mimic residual uncertainty in the predictions with randomly generated variation; add DBH at the beginning of the year, predicted annual DBH growth, and residual variation to obtain DBH for the beginning of the subsequent year; update BA and BAL for these plots for the current year; record the year, tree DBHs, and plot basal area per acres estimates;

5. Use plot-level estimates from Phase 2, Steps 4 and 5, and calculate the mean and variance of basal area per acre across all plots; record the year, mean, and variance of the basal area per acre estimates;

6. Repeat Phase 1, Steps 1-6 until Year 20 has been reached;

7. Repeat Phase 1 and 2 a large number of times.

For each year, "true" (Phase 1) means and 95% confidence intervals and Phase 2 means and 95% confidence intervals were calculated for basal area per acre for each estimation alternative to obtain comparisons of bias and precision and to assess the effects of imprecision in the model diameter growth predictions on the imprecision of the basal area per acre estimates. In addition, estimates of precision for each tree and plot can be obtained from the distributions of DBH and basal area per acre estimates, respectively, obtained from Phase 2.

The results are summarized in graphs of the distributions of coefficients of variation after 5 years for predicted tree DBH (Figure 1), predicted plot basal area per acre (Figure 2), and means and 95% confidence intervals for annual basal area per acre estimates across all plots for each estimation alternative compared to the "true" estimates (Figures 3,4,5).

Figure 1. Coefficients of variation for 5-year individual tree DBH predictions.

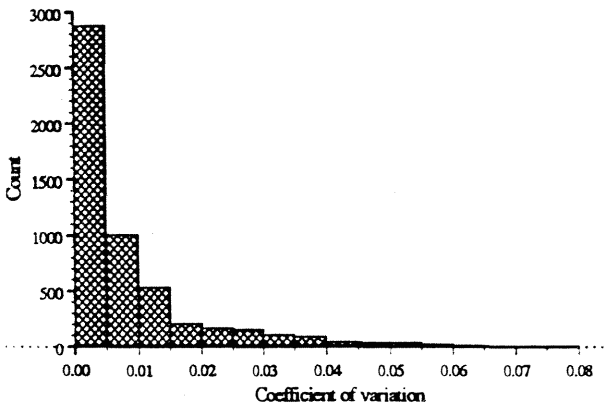
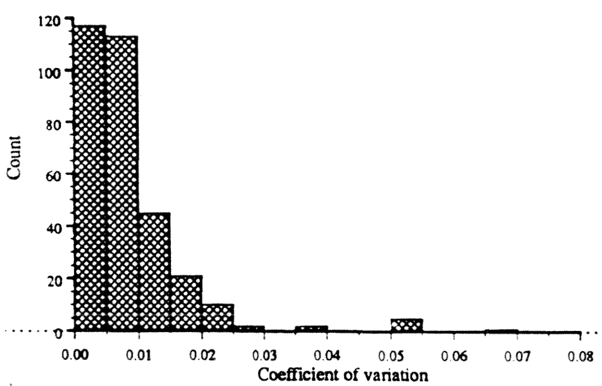


Figure 2. Coefficients of variation for 5-year plot basal area predictions.



RESULTS AND CONCLUSIONS

Neither individual tree DBH growth nor plot basal area growth per acre was well-estimated. Coefficients of variation for 5-year DBH growth predictions ranged from 0.2 to 1.2, while coefficients of variation for basal area growth per acre ranged from 0.1 to 0.6. The smaller values for basal area growth are attributed to the effects of aggregating tree DBH predictions over plots.

Nevertheless, 5-year predictions of tree DBH and plot basal area were well-estimated with coefficients of variation for both below 0.1 with most below 0.03 (Figures 1,2). These results are attributed to the

Figure 3. Mean basal area per acre estimates and 95% confidence intervals for "true" data and 20% annual samples.

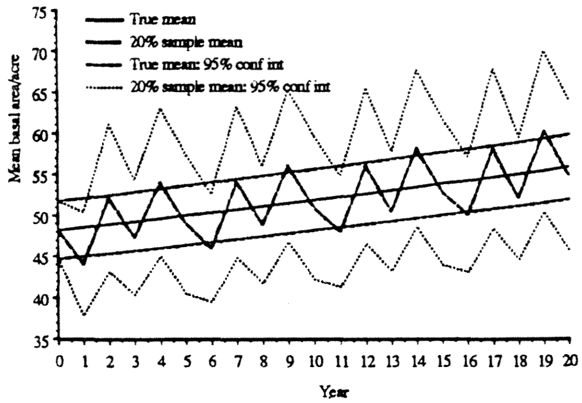


Figure 4. Mean basal area per acre estimates and 95% confidence intervals for "true" data and 5-year moving average.

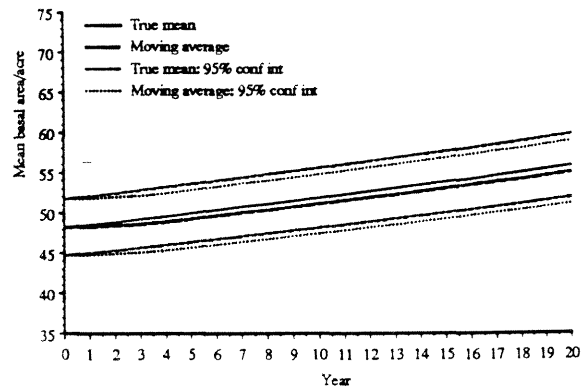
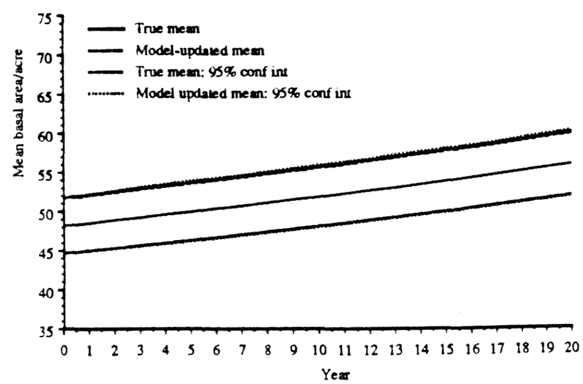


Figure 5. Mean basal area per acre estimates and 95% confidence intervals for "true" data and model updates.



observation that the previous 5-year growth typically makes only a small contribution to size for all but the smallest trees. Thus, despite considerable uncertainty in the growth predictions, knowledge of size at the beginning of the 5-year interval permits quite precise estimation of size at the end of the period.

The monotonic increasing form of the "true" means from Phase 1 is an artifact of the simulation procedure which made no provision for regeneration, mortality, removals, or land use change in either the population or the modeling. A complete analysis requires unbiased models for these components also. Nevertheless, the relatively large standard error of the mean basal area per acre estimates reflects the great variability that exists naturally among plots. This uncertainty is accentuated for the annual 20% samples (Figure 3) whose annual means sometimes fall outside the 95% confidence interval for the "true" means.

The observed lag of the 5-year moving averages from the "true" means is at least partially an artifact of the simulation procedure which permits only a monotonically increasing trend in basal area. However, any multi-year moving average should be expected to be offset from "true" mean due to the nature of the estimate. This offset is the primary detractor for the moving average approach, because the precision estimates for this approach are quite accurate (Figure 4).

The virtual coincidence of the "true" means and model-updated means is also at least partially dependent on the use of the simulated "true" data to mimic observed growth. The important finding for this approach is that the uncertainty associated with the growth model predictions contributes very little to the total uncertainty of the basal area per acre means (Figure 5). This result is attributed to two previously noted effects: first, the excellent precision of the plot basal area per acre estimates, and second, the relatively large natural variation among plots.

The overall conclusion of the study is that despite the relatively large uncertainty in the growth predictions, the model updating approach produces mean basal area per acre means that are nearly as precise as the "true" means but without the lag or offset effect of the moving average approach.

REFERENCES

Eagar, C., Miller-Weeks, M., Gillespie, A.J.R., and Burkman, W. (1991), "Summary Report: Forest Health Monitoring - New England/Mid-Atlantic," NE-INF-115-92, U.S. Department of Agriculture, Forest Service, Radnor, PA.

Holdaway, M.R. (1999), "An FIA Tree Growth Model for Updating Annual Forest Inventories in Minnesota," in Integrating Tools for Natural Resources Inventories in the 21st Century, eds. M. Hansen, S. Fairweather, and T. Burk, General Technical Report, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN.

Linhart, H., and Zucchini, W. 1986. Model selection. John Wiley and Sons, Inc. New York.

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

McRoberts, R.E. 1998. Comparing approaches to calibrating an annual growth model using non-annual size observations. American Statistical Association 1998 Proceedings of the Biometrics Section. Alexandria, VA.

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

USDA, Forest Service (1970), Operational Procedures, Forest Service Handbook 4809.11, Chapter 10: 11.1–1 - 11.1–3, USDA Forest Service, Washington, DC.

White, D., Kimerling, J., and Overton, S.W. (1992), Cartographic and Geometric Components of a Global Sampling Design for Environmental Monitoring, Cartographic and Geographic Information Sciences, 19, 5-22.