Imputation and Model-Based Updating Techniques for Annual Forest Inventories

Ronald E. McRoberts

ABSTRACT. The USDA Forest Service is developing an annual inventory system to establish the capability of producing annual estimates of timber volume and related variables. The inventory system features measurement of an annual sample of field plots with options for updating data for plots measured in previous years. One imputation and two model-based updating techniques are described and evaluated with respect to the bias and precision of their annual estimates of basal area per unit area. The evaluations indicate that simple plot-level imputation and model-based updating techniques produce adequately unbiased and precise estimates of annual mean basal area per unit area for large areas. For. Sci. 47(3):322–330.

Key Words: Growth models, remote sensing, simulation.

The U.S. Renewable Forest and Rangeland Resources Planning Act of 1978 requires that the USDA Forest Service conduct inventories of forestland to determine its extent and condition and the volume of standing timber, timber growth, and timber removals. The U.S. Agricultural Research Extension and Education Reform Act of 1998 further requires that the Forest Service conduct annual forest inventories in all states, measuring 20% of the plots in each state each year.

Forest Inventory and Analysis (FIA) precision standards (USDA FS 1970) require a sampling intensity of 1 plot for approximately every 2,428 ha. To satisfy this requirement, the geographical sampling hexagons established for the Forest Health Monitoring Program (White et al. 1992) were divided into 27 smaller FIA hexagons, each containing approximately 2,403 ha. An equal probability sample of field plots, designated the federal base sample, was constructed by establishing a plot in each FIA hexagon (Brand et al. 2000). The federal base sample was systematically divided into five interpenetrating, nonoverlapping panels (Figure 1). Each year, the plots in a single panel are selected for measurement, and panels are selected on a 5 yr, rotating basis.

At least three approaches to calculating annual FIA estimates using data acquired from the federal base sample have been considered (McRoberts 1999). The simplest approach is to use the data from the 20% panel of plots measured in the current year. Although the resulting estimates would reflect current conditions, their low precision may be unacceptable for some variables due to the small annual sample size. A second approach is to use the most recent data for all plots, i.e., the five most recent panels of measurements, and employ a moving average estimator. The advantage of this approach is that precision is increased because data for all plots are used for estimation; the disadvantage is that the estimates reflect a moving average of conditions over the past 5 yr and may seriously trail current conditions in the presence of time-based trends. A third approach is to update to the current year data for plots measured in previous years and then base estimates on the data for all plots. If the updating procedure is unbiased and sufficiently precise, this approach provides nearly the same precision as using all plots but without the adverse effects of using out-of-date information.

The objective of the study was to demonstrate that an approach to annual inventory estimation that uses simple updating techniques is a feasible, unbiased, and precise alternative; the objective was not necessarily to develop updating techniques with the greatest optimality, precision, or explanatory power. Two categories of updating

Ronald E. McRoberts, Mathematical Statistician, North Central Research Station, USDA Forest Service, 1992 Folwell Avenue, St. Paul, MN 55108—Phone: (651) 649-5174; Fax: (651) 649-5285. E-mail: rmcroberts@fs.fed.us.

Manuscript received February 11, 2000. Accepted April 3, 2001.

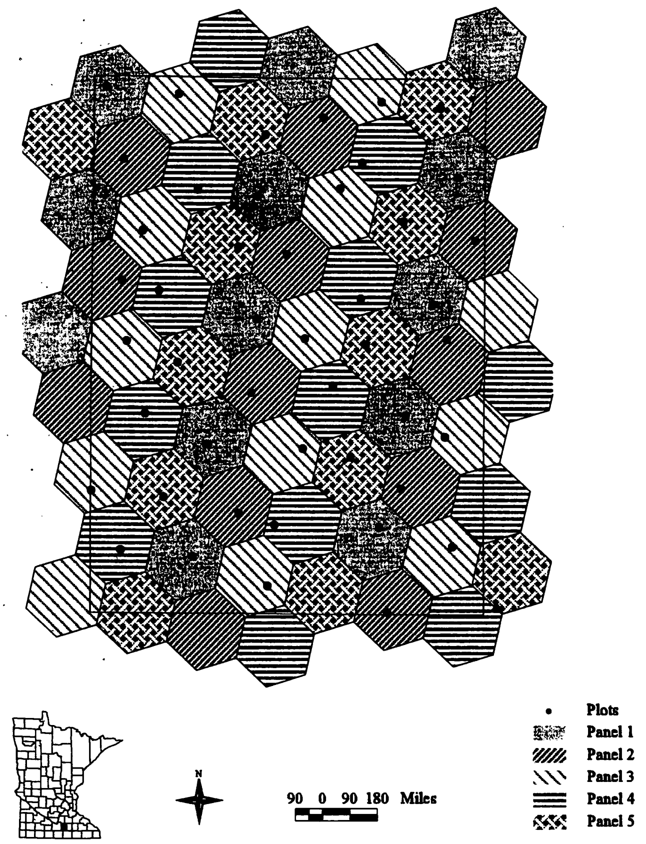

Figure 1. The hexagon-based, interpenetrating FIA sampling design for Waseca County, Minnesota.

techniques, imputation and modeling, are of general interest and were evaluated with respect to the quality of basal area per unit area estimates using a database of annual tree information. Although basal area is only one of several inventory variables of interest, similar results would be expected for variables highly correlated with basal area such as volume and stems per unit area.

Annual Database

The database of annual tree information was constructed using individual tree observations obtained from the 5,287 timberland plots measured in both the 1977 (Spencer 1982) and 1990 (Miles et al. 1995) USDA Forest Service periodic inventories of Minnesota. These plots represented approximately 5.95 million ha of timberland, defined as forestland capable of producing at least 1.40 m^3/ha/yr of industrial wood crops under natural conditions (Miles et al.1995). Plots included in the 1977 inventory were actually measured between 1974 and 1978; plots included in the 1990 inventory were actually measured between 1986 and 1991. Thus, the time interval between measurements for some plots was less than 10 yr. Because the intent was to construct a database consisting of 10 yr of growth, plots with remeasurement intervals of less than 10 yr were excluded, leaving 4,855 plots on which there were 100,897 trees.

Plots measured for the 1977 and 1990 Minnesota inventories consisted of 10 subplots described as variable radius plots due to the use of point sampling techniques. With these techniques, trees are selected with probability proportional to cross-sectional area rather than proportional to the frequency of occurrence in the population (Myers and Beers 1971). With point sampling, the number of trees in the population represented by a sample tree, termed the *tree factor,* is calculated as a scaling constant divided by the square of the tree diameter. Tree factors are used to expand the measured attributes of sample trees to per unit area estimates.

Based on observations of individual trees with diameters at breast height (dbh) (1.37 m above ground) of at least 12.7 cm, an 11 yr database was constructed. The database consisted of annual dbh and annual survival, ingrowth, mortality, or harvest status of each tree. To construct the database, total growth between inventories was distributed over varying numbers of years for individual trees in each of four categories: (1) survivor trees alive at both inventories, (2) ingrowth trees that attained the 12.7 cm minimum dbh between inventories, (3) mortality trees that died between inventories due to causes other than harvest, and (4) harvest trees that were harvested between inventories. For each survivor tree, average annual dbh growth was calculated by dividing the total dbh growth over the measurement interval by the number of years between measurements. Measured dbh for the 1977 inventory was assigned to year 0, and dbh's for the 10 subsequent years were calculated by adding the average annual growth to the previous year's dbh. Because ingrowth trees were measured only in the 1990 inventory, dbh measurements for those trees were assigned to year 10, and dbh's for previous years back to year 0 were sequentially estimated by subtracting predictions of annual dbh growth obtained from individual tree growth models (Lessard et al. 2001) from the current year's dbh. Ingrowth status for these trees was designated in the year the tree attained the 12.7 cm minimum dbh. For trees that died due to causes other than harvest, a year of mortality between 1 and 10 was randomly selected and assigned to the tree independently of years of mortality assigned to other trees on the plot. For harvested trees, a year of harvest was randomly selected and assigned to all harvested trees on the plot. For both mortality and harvest trees, measured dbh for the 1977 inventory was assigned to year 0, and dbh's for subsequent years up to the year of mortality or harvest were calculated by adding previous year's dbh and predictions of annual growth obtained from the individual tree diameter growth models. Although these procedures create greater uniformity in annual dbh growth than would be observed, the effects of differences between actual and calculated growth are expected to have minimal impact on evaluations of the updating techniques. Alternatives would require either annual remeasurement or destructive sampling of all trees.

Evaluations of the updating techniques were based on plot basal area per hectare (BA), a variable representing the sum, scaled to a per hectare basis using tree factors, of the cross-sectional areas of live tree boles at breast height. Calculation of unbiased estimates of annual change in basal area per hectare (ΔBA) is difficult using data from variable radius plots (Van Deusen et al. 1986). For this evaluation, tree factors corresponding to dbh's for year 0 were calculated for all trees and then held constant for the entire 10 yr interval.

With the database of annual tree diameter values and these tree factors, BA was calculated each year for each plot, and ΔBA was calculated each year for each plot as the difference between BA for the current and previous years.

Updating Techniques

Although the use of updating techniques in forest inventory is not well documented, it is not a new phenomenon. At the Southern Research Station (SRS) of the USDA Forest Service, data for inaccessible plots have been replaced or imputed by substituting either data for a plot selected from the same stratum or the stratum mean (Reams and Van Deusen 1999). At the North Central Research Station (NCRS) of the USDA Forest Service, data for a proportion of well-established, undisturbed FIA plots have been updated using the STEMS (Belcher et al. 1982) individual tree survival and diameter growth models (Miles et al. 1995). Whereas the basis of the SRS updating techniques has been current, whole-plot quantities, the basis of the NCRS updating techniques has been change in tree conditions since the last inventory. NCRS envisions model-based updating as an integral component of annual inventory systems and is developing a new set of survival and growth models specifically designed for this purpose (McRoberts and Hansen 1999, Lessard et al. submitted).

The basis for updating for this study was plot-level change. Change, rather than current condition, was selected, because plot BA at an initial time accounts for a very large proportion of plot BA 5 yr later, except for newly regenerated plots. Therefore, any bias or imprecision in the updating technique affects only a small proportion of a 5 yr BA prediction. Updating at the plot level, as opposed to the tree level, was investigated to examine the feasibility, bias, and precision of simpler procedures. One imputation and two model-based updating procedures were investigated; all three were based on the assumption that ΔBA is related to BA at the end of the previous year.

The IMPUTE Approach

An approach to dealing with plots that have been measured in a previous year, but not the current year, is to consider their current year observations as missing. Imputation techniques address the missing observation problem by seeking plausible and consistent replacement observations. Among the imputation techniques Sande (1982) describes for selecting replacement observations, hotdeck procedures select from a pool of current observations that match prescribed attributes of the missing observations, and nearest neighbor procedures select from a pool of current observations that are only similar to the missing observations with respect to the prescribed attributes. Rubin (1978) advocates multiple completions of the data sets via imputation of missing observations to allow assessment of the uncertainty in imputed variables and to protect against extreme results. Van Deusen (1997) and Reams and Van Deusen (1999) proposed using the imputation techniques discussed by Sande (1982, 1983), Rubin and Schenker (1986), and Rubin (1987) for replacing missing inventory observations. Reams and McCollum (2000) reported on the use of hotdeck imputation for replacing missing observations of plot-level merchantable volume per unit area.

For this investigation, a variation of the single-imputation, nearest neighbor approach was used to update BA for plots measured in previous years. The procedure consisted of three steps: (1) plots measured in the current year were placed into similarity groups with respect to previous year's BA; (2) each plot measured in a previous year was matched to a group of similar plots measured in the current year; and (3) a ΔBA value from the group of similar plots measured in the current year was selected to replace the missing observation for the plot measured in the previous year. The groups were created by first ordering all plots measured in the current year with respect to previous year's BA and then creating groups of 20 consecutive plots beginning with the plot with the lowest previous year's BA. Individual plots measured in previous years were then matched to a group of plots measured in the current year on the basis of previous year's BA, whether previous year's BA was obtained as a measurement or as an updated estimate. For each plot measured in a previous year, a plot was randomly selected with replacement from the group of 20 similar plots, and the latter plot's average ΔBA since last measurement was imputed as current ΔBA to the former plot. This imputation approach used 5 yr average ΔBA as a surrogate for current ΔBA, was expected to produce unbiased estimates of BA means across all plots, and is hereafter described as IMPUTE.

The PREDICT Approach

Two model-based updating techniques were also investigated. For each modeling technique, ΔBA for a plot was assumed to be related to both BA at the end of the previous year and to the current survival, ingrowth, mortality, or harvest status of trees on the plot.

For the first model-based updating technique, the annual status of trees was used to place all plots into one of three categories: (1) survival: no mortality or harvested trees; (2) mortality: at least one mortality tree, but no harvested trees; and (3) harvest: at least one harvested tree. Because the annual survival, mortality, and harvest status of plots will be known only for measured plots, a technique was developed for predicting the annual status of plots measured in previous years with respect to these three categories. First, all plots in the annual database were ordered by previous year's BA and then placed into groups of 250 consecutive plots beginning with the plot with the lowest previous year's BA. For the kth group, the proportions of plots, $P_{k,surv}$, $P_{k,mort}$, and $P_{k,harv}$, in the survival, mortality, and harvest categories, respectively, were calculated, and models were formulated:

$$E(P_{k,surv}) = \exp(\beta_1 \mathrm{BA}_k^{\beta_2}), \tag{1a}$$

$$E(P_{k,mort}) = \exp(\beta_3 \mathrm{BA}_k^{\beta_4}), \tag{1b}$$

and

$$E(P_{k,harv}) = 1 - E(P_{k,surv}) - E(P_{k,mort}), \tag{1c}$$

where $E(.)$ denotes statistical expectation and BA_k denotes the BA mean for the kth group. The parameters, β, of [1a–c] were estimated by maximizing the likelihood,

$$L = \prod_k \left(\frac{N_k}{n_{k,surv}\; n_{k,mort}\; n_{k,harv}} \right) P_{k,surv}{}^{n_{k,surv}} P_{k,mort}{}^{n_{k,mort}} P_{k,harv}{}^{n_{k,harv}} \tag{2}$$

where $N_k = 250$ is the total number of plots in the kth group and $n_{k,surv}$, $n_{k,mort}$, and $n_{k,harv}$ are the numbers of survival, mortality, and harvest plots, respectively.

For each of the three categories of plots, a model of the relationship between ΔBA and previous year's BA was formulated, and the model parameters were estimated using regression techniques. Graphs of ΔBA versus previous year's BA for the survival (Figure 2a) and mortality plots revealed a series of bands of observations that corresponded to ΔBA for the same net change in number of plot trees. This effect is attributed to the observation that the contribution to ΔBA of a single ingrowth or mortality tree typically was much greater than the combined contributions to ΔBA of growth of surviving trees. Because sampling on the 10 variable radius subplots was conducted using an 8.61 m^2/ha basal area factor (Spencer 1982, Miles et al. 1995), the absolute value of the contribution to ΔBA of a single ingrowth or mortality tree is 0.86 m^2/ha in year 0 and slightly more in subsequent years because of growth of the tree (Myers and Beers 1971). Therefore, to simplify the modeling approach, the net change in number of trees for survival and mortality plots was determined and multiplied by 0.86 m^2/ha, and the product was subtracted from the net ΔBA for the plot. The relationships between the resulting annual standardized change in basal area, ΔBA_{sta} and previous year's BA were graphed (Figure 2b), simple models of the relationships were formulated, and the model parameters were estimated—for the survival category,

$$E(\Delta BA_{sta}) = \beta_1[1 - \exp(\beta_2 BA)], \tag{3a}$$

and for the mortality category,

$$E(\Delta BA_{sta}) = \beta_1 + \beta_2 BA + \beta_2 BA^2, \tag{3b}$$

where $E(.)$ is statistical expectation, and the β's are parameters that were estimated using weighted nonlinear regression techniques.

Models of the relationships between net change in number of plot trees and previous year's BA were also formulated for the survival and mortality categories. For the kth group, the model for the probability, $P_{k,surv,j}$, of j ingrowth trees for survival plots was formulated as,

$$E(P_{k,surv,j}) = \beta_{j1} + \beta_{j2} BA_k + \beta_{j3} BA_k{}^2, \tag{4a}$$

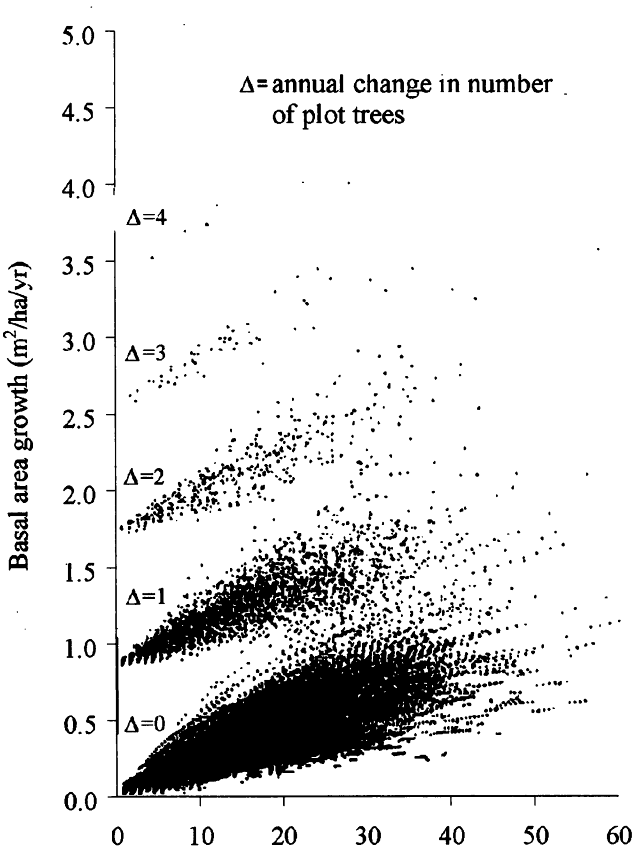

Figure 2a. Observed basal area growth versus basal area for plots with no mortality or harvested trees (includes trees entering the inventory).

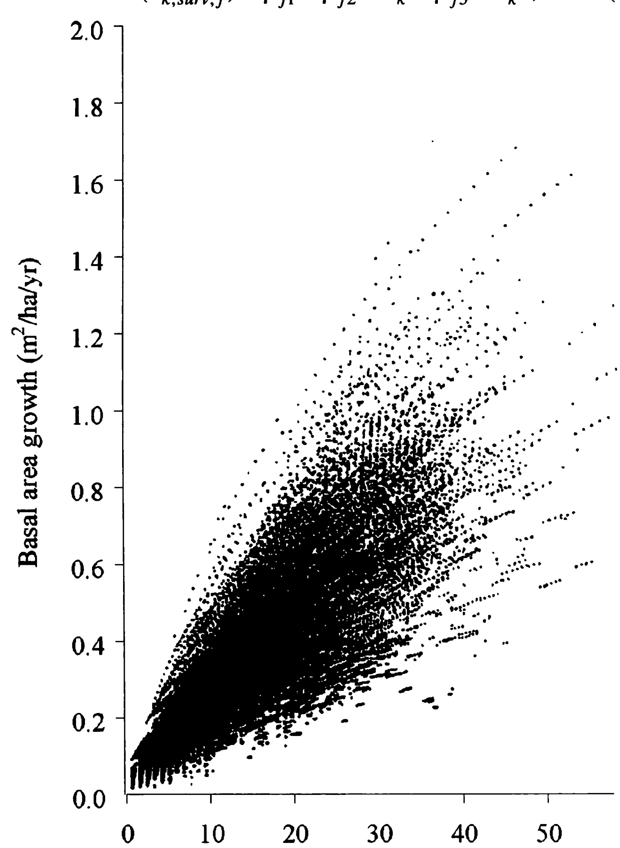

Figure 2b. Standardized basal area growth versus basal are[a] plots with no mortality or harvested trees. Standardized b[asal] area growth is observed basal area growth minus the produ[ct] annual change in number of plot trees and the basal area fa[ctor]

for $j = 0,1,2$, or 3. Because the maximum number of ingrowth trees never exceeded four in the observed data, the model for $P_{k,surv,4}$ was formulated as

$$E(P_{k,surv,4}) = 1 - E(P_{k,surv,0}) - E(P_{k,surv,1}) - E(P_{k,surv,2}) - E(P_{k,surv,3}) \tag{4b}$$

The parameters of (4a) were estimated by maximizing the likelihood,

$$L = \prod_k \left(\frac{N_k}{n_{k,surv,0} \ldots n_{k,surv,4}} \right) P_{k,surv,0}{}^{n_{k,surv,0}} \cdots P_{k,surv,4}{}^{n_{k,surv,4}} \tag{5}$$

where $N_k = 250$ is the total number of plots in the kth group and $n_{k,surv,0}$, $n_{k,surv,1}$, $n_{k,surv,2}$, $n_{k,surv,3}$, and $n_{k,surv,4}$ are the numbers of survival plots with zero, one, two, three, and four ingrowth trees, respectively. For the mortality category, similar models and techniques were used for estimating the net annual change in number of plot trees that ranged from zero to four in the observed data.

For the harvest category, the relationship between ΔBA and previous year's BA was estimated directly using a simple linear model,

$$E(\Delta BA) = \beta_1 + \beta_2 BA \tag{6}$$

with the parameters estimated using weighted linear regression techniques.

Graphs of the relationships between both ΔBA and ΔBA_{sta} and previous year's BA consistently exhibited heterogeneity of variance, which necessitated weighted regression approaches to parameter estimation. To obtain weights, standard deviations were estimated for the distributions of residuals for narrow ranges of predicted ΔBA, and relationships between $\hat{\sigma}$ and $\Delta \hat{BA}$ were described with simple linear models of the form:

$$\hat{\sigma} = \alpha_1 + \alpha_2 \Delta \hat{BA} \tag{7}$$

These relationships provided the basis for selecting weights for weighted regressions and were used to select random residuals for incorporating uncertainty into ΔBA predictions.

With this updating technique, hereafter described as PREDICT, the procedure for predicting ΔBA for each plot measured in a previous year is summarized as follows:

1. Create plot groups:
2. Order plots measured in previous years by BA for the year immediately preceding the current year;
3. Create groups of 20 consecutive plots beginning with the plot with lowest BA for the year immediately preceding the current year.
4. Predict plot survival, mortality, or harvest status: Randomly select a number, x, from a Uniform (0,1) distribution;
   a. If $0 \le x < E(P_{k,surv})$, predict survival;
   b. If $E(P_{k,surv}) \le x < E(P_{k,surv}) + E(P_{k,mort})$, predict mortality;
   c. If $E(P_{k,surv}) + E(P_{k,mort}) \le x$, predict harvest.
3. Predict ΔBA:
   a. Survival plots: Predict ΔBA as the sum of three components:
      (i) expected standardized ΔBA from (3a);
      (ii) a residual calculated as the product of a number randomly selected from a Normal (0,1) distribution and a standard deviation calculated from (7);
      (iii) the product of 0.86 m^2/ha and a prediction for the number of ingrowth trees obtained using (4a-b), a number randomly selected from a Uniform (0,1) distribution, and a procedure similar to that for selecting the survival, mortality, or harvest status of plots;
   b. Mortality plots: Predict ΔBA in a manner similar to that for survival plots;
   c. Harvest plots: Predict ΔBA as the sum of expected ΔBA from [6] and a residual selected in the same manner as for survival plots.

The REMOTE Approach

Although model predictions of survival, mortality, and harvest status were expected to be unbiased, the combined effects of their uncertainties and those of the ΔBA predictions may increase the estimated standard errors of the annual BA means, the variability of the estimates of the means around the estimates based on a complete sample, or both. Thus, a second model updating technique was developed based on the assumption that satellite-based remote sensing techniques can be used to detect plots that have experienced substantial disturbance (Befort 2000). Disturbance for this technique may be due to either mortality or harvest; no distinction is made. Remote sensing techniques are assumed to be adequate for confidently detecting disturbed plots satisfying two criteria: previous year's BA ≥ 2.79 m^2/ha, and (ΔBA/BA) ≤ –0.2 (Befort, pers. comm.).[1]

For plots satisfying these criteria, ΔBA was predicted as the product of previous year's BA and the ratio of ΔBA/BA in the annual database. For plots not satisfying these criteria, a model for ΔBA was formulated as

$$E(\Delta BA) = \beta_1 + \beta_2[1 - \exp(\beta_3 BA)] \tag{8}$$

and its parameters were estimated using weighted nonlinear regression techniques.

With this approach, hereafter described as REMOTE, updating again involves prediction of both status and ΔBA. First, plots measured in previous years that satisfy the remote sensing disturbance detection criteria are identified, and their ΔBA predictions are obtained as the product of previous year's BA and the ratio of ΔBA/BA in the annual database. Second, for the remaining plots, ΔBA is predicted as the sum of expected ΔBA obtained from (8) and a residual obtained as

[1] William Befort, Division of Forestry, Minnesota Department of Natural Resources, November 9, 1999.

the product of a number randomly selected from a Normal (0,1) distribution and a standard deviation calculated from (7). An important feature of the REMOTE technique is that the status and ΔBA for plots with the greatest loss of BA are predicted using simulated remote sensing techniques, not models.

Simulating the Inventory

The feasibility of the updating techniques and the bias and precision of their estimates of annual BA means were evaluated by simulating annual inventory procedures for the 5.95 million timberland ha in Minnesota. To mimic the annual inventory intensity of one plot per 2,403 ha, 2,476 plots were randomly selected for each simulation from among the 4,855 timberland plots. The 2,476 selected plots were ordered by their FIA plot numbers and distributed among five equal-sized panels by systematically assigning every fifth plot to the same panel. Because FIA plot numbers had been assigned sequentially on the basis of their geographic locations, the panel assignments for each simulation approximated the systematic, interpenetrating feature of the annual inventory design. Each simulation was initiated with a complete inventory of the 2,476 plots, simulated by beginning with the annual database year 0 values for each plot. On a rotating basis, one 20% panel of plots was selected for measurement each year. Simulated measurement of a plot consisted of replacing its estimated BA value with the value for the appropriate year in the annual database. For the remaining 80% of plots, plot data were updated using each of the three updating techniques. Each year, estimates of BA means across all plots and the standard errors of the means were calculated for each updating technique. In addition, each year estimates of the current BA mean and standard error of the mean were obtained using three additional estimation methods: (1) the SAMPLE100 estimates were based on the annual database values for all 2,476 plots selected for the simulation; (2) the SAMPLE20 estimates were based on the annual database values for the 20% annual sample; and (3) the MOVING estimates were based on the most recent panel of measurements for all plots. The SAMPLE100 estimates are the standard for comparisons because they are based on measured, not updated, annual BA for 100% of the plots selected for the particular simulation. Although 250 simulations were conducted, median values for annual BA for individual plots, annual BA means over all plots, and the standard errors of the annual BA means stabilized for all estimation methods by approximately 175 simulations.

The performance of the updating techniques was evaluated for smaller areas by simulating the annual inventory process for two individual counties. St. Louis County in northeastern Minnesota in the Aspen-Birch FIA unit includes approximately 1.1 million ha of timberland; Pine County in east central Minnesota in the Central Hardwood FIA unit includes approximately 0.2 million ha of timberland. To mimic the annual inventory intensity of one plot per 2,403 ha, 455 of the 877 St. Louis County timberland plots were randomly selected for each simulation, while 90 of the 172 Pine County timberland plots were randomly selected for each simulation. Source data for the imputation updating technique were obtained from two different pools of plot data; the first pool consisted of the 20% annual sample of the 2,476 plots selected for the entire state, while the second pool was restricted to the 91 plots for St. Louis County and the 18 plots for Pine County. Because of these small annual samples the number of plots in similarity groupings for the second pools were reduced from 20 as used for the statewide analyses to 10 for St. Louis County and to 5 for Pine County. For the modeling technique, the same models used for the statewide analyses were used for these county-level analyses.

Analyses

Only graphical and tabular results are shown for estimates obtained with the SAMPLE20 and MOVING techniques; n tests of significance were performed. The intent is merely illustrate the undesirable features of these estimates and th motivation for developing updating techniques. Th SAMPLE20 estimates are based on sample sizes that a known to be inadequate for satisfying national precisio standards,while the MOVING estimates are necessari biased whenever there is a time-based trend in the unde lying data.

For each updating technique, the median of the distr butions of the 250 simulated estimates of the annual B means and the medians of the distributions of the 2 simulated estimates of the standard errors of the mea were determined. The nonparametric Wilcoxon Sign Ranks Test (Conover 1980, p. 280–292) was used compare the updating techniques with respect to the es mates of the annual BA means they produced. The basis the comparisons is annual differences between SAMPLE100 estimates and the medians of the distri tions of the estimates of annual BA means for the updat techniques. The same test was used to compare SAMPLE100 estimates of the standard errors of the me and the medians of the distributions of the correspond estimates obtained with the three updating techniques

Results

Results of comparing the SAMPLE20 and MOVI medians of statewide estimates of annual BA means to statewide SAMPLE100 estimates were consistent with pectations (Table 1, Figure 3a). First, the SAMPLE20 m ans exhibited considerable variation around the SAMPLE estimates because of their smaller sample sizes; second, s SAMPLE20 medians fell outside the 2-standard error i val around the SAMPLE100 estimates; third, the MOV medians exhibited systematic bias, a result of the trend i underlying data as revealed by the SAMPLE100 estim and fourth, the SAMPLE20 medians of the estimates c standard errors of the means were much larger thai SAMPLE100 estimates. Because of these undesirable and precision characteristics of estimates obtained wit SAMPLE20 and MOVING techniques, the remaining cussion focuses exclusively on comparing results obt with the updating techniques to the SAMPLE100 estin

Table 1. Median values for 250 simulations of Minnesota timberland estimates of annual BA mean (m^3/ha) and standard error of the mean (m^3/ha).

	SAMPLE100		SAMPLE20		MOVING		IMPUTE		PREDICT		REMOTE	
Year	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	16.52	0.14	16.50	0.30	16.56	0.14	16.49	0.14	16.48	0.14	16.51	0.14
2	16.28	0.14	16.43	0.31	16.51	0.14	16.37	0.14	16.37	0.14	16.34	0.14
3	16.17	0.14	16.43	0.31	16.44	0.14	16.25	0.14	16.27	0.14	16.23	0.14
4	16.12	0.14	16.24	0.31	16.38	0.14	16.18	0.14	16.19	0.14	16.16	0.14
5	16.14	0.14	15.62	0.32	16.25	0.14	16.03	0.14	16.05	0.14	16.06	0.14
6	16.07	0.14	15.83	0.32	16.11	0.14	15.89	0.14	15.92	0.14	15.97	0.14
7	15.87	0.14	15.94	0.32	16.01	0.14	15.81	0.15	15.82	0.14	15.87	0.14
8	15.83	0.15	16.18	0.33	15.96	0.14	15.81	0.15	15.78	0.14	15.83	0.14
9	15.84	0.15	16.24	0.34	15.96	0.14	15.89	0.15	15.78	0.14	15.76	0.15
10	15.63	0.15	15.08	0.34	15.85	0.15	15.75	0.15	15.67	0.15	15.62	0.15

Medians of statewide estimates produced using the updating techniques were not statistically significantly (α = 0.05) different from the SAMPLE100 estimates, either for annual BA means or for the standard errors of the means (Table 1). In addition, pairwise comparisons of the updating techniques did not reveal statistically significant (α = 0.05) differences among them with respect to the medians of statewide estimates of annual BA means or medians of the estimates of standard errors of the means. Nevertheless, the medians of distributions of estimates produced by the REMOTE technique were consistently closer to the SAMPLE100 estimates with respect to both mean deviation and root mean squared deviation. Graphs of the SAMPLE100 estimates and medians of the IMPUTE, PREDICT, and REMOTE estimates of annual BA means confirm the results of the significance testing that the medians obtained with the updating techniques closely follow the SAMPLE100 estimates, and that the differences among them are small (Figure 3b).

For St. Louis County, the large, heavily forested county with many plots, no statistically significant (α = 0.05) differences in estimates of annual BA means were detected when the IMPUTE medians obtained using statewide similarity groups, the PREDICT medians, and the REMOTE medians were compared to the SAMPLE100 estimates (Table 2). However, the IMPUTE medians obtained using county similarity groups were statistically significantly (α = 0.05) different from the SAMPLE100 estimates.

For Pine County, the smaller county with fewer plots, the median IMPUTE estimates of annual BA means obtained with both statewide and county similarity groups and the median PREDICT estimates were statistically significantly (α = 0.05)) greater than the SAMPLE100 estimates (Table 3). No statistically significant (α = 0.05) differences were detected between the REMOTE medians and the SAMPLE100 estimates. Also, no statistically significant (α = 0.05) differences were detected between

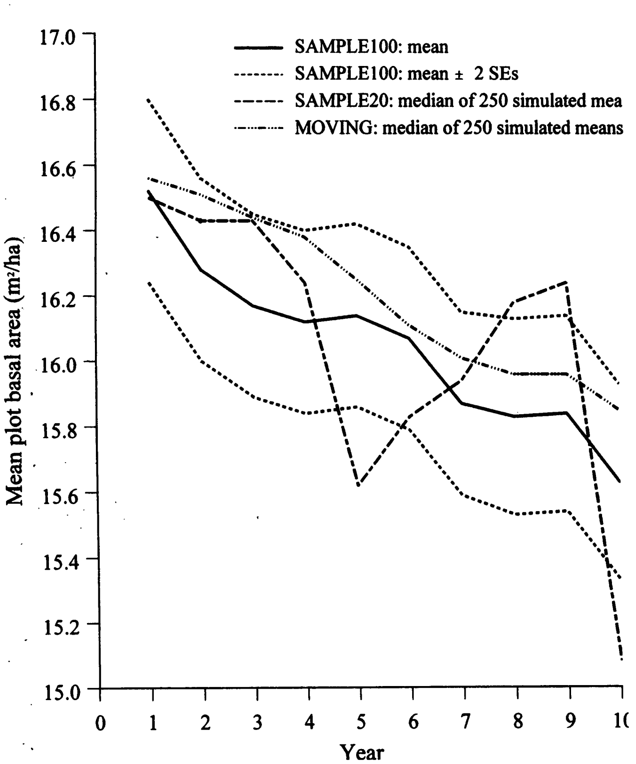

Figure 3a. Mean plot basal area estimates.

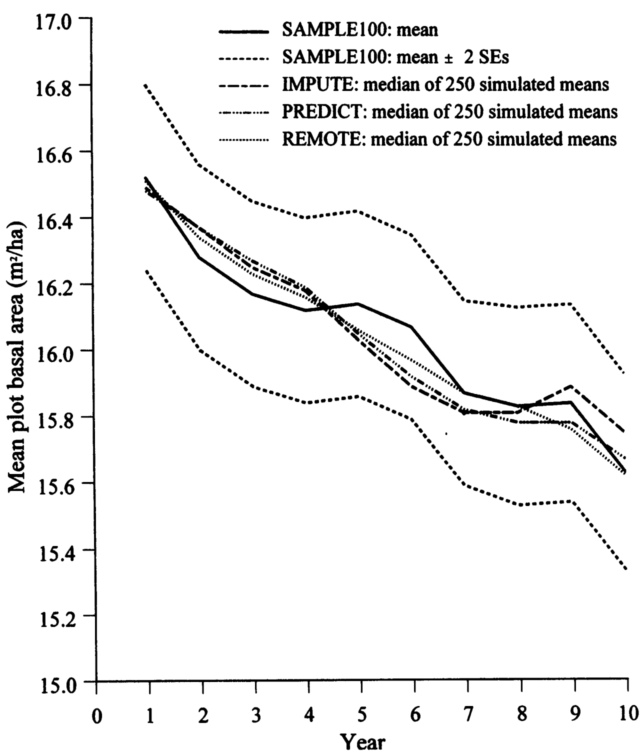

Figure 3b. Updated mean plot basal area estimates.

Table 2. Median values for 250 simulations of St. Louis County, Minnesota, timberland estimates of annual mean BA (m^3/ha) and standard error of the mean (m^3/ha).

			IMPUTE							
	SAMPLE100		Statewide		County		PREDICT		REMOTE	
Year	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	14.87	0.30	14,81	0.30	14.85	0.30	14.82	0.29	14.86	0.30
2	14.58	0.30	14.68	0.31	14.83	0.30	14.77	0.30	14.65	0.30
3	14.48	0.30	14.55	0.31	14.62	0.31	14.68	0.30	14.56	0.30
4	14.41	0.31	14.51	0.31	14.58	0.31	14.62	0.30	14.46	0.30
5	14.46	0.31	14.27	0.31	14.28	0.31	14.39	0.30	14.41	0.31
6	14.42	0.31	14.22	0.31	14.30	0.31	14.34	0.31	14.33	0.31
7	14.23	0.32	14.13	0.32	14.30	0.31	14.21	0.31	14.25	0.31
8	14.26	0.32	14.09	0.32	14.05	0.32	14.17	0.31	14.29	0.32
9	14.37	0.33	14.20	0.33	14.18	0.33	14.21	0.32	14.30	0.32
10	14.26	0.33	14.12	0.33	14.13	0.33	14.10	0.32	14.28	0.33

the medians of the estimates of standard errors of the means for any of the updating techniques and the SAMPLE100 estimates.

Conclusions

Several conclusions may be drawn from these analyses. First, the quality of estimated annual BA means and standard errors of the means obtained with the imputation and model-based updating techniques is generally similar for both the statewide and county estimates. The single exception is that the large county IMPUTE medians obtained using statewide similarity groups were significantly greater than the SAMPLE100 estimates. For the small county, only the REMOTE technique produced estimates with medians that were not statistically significantly different from the SAMPLE100 estimates. This result suggests that estimation for individual counties with small numbers of plots should be approached with caution with any technique.

The second conclusion is that at both state and county levels, all three updating techniques produced estimates of the standard errors of the means that did not differ significantly from the SAMPLE100 estimates and that for practical applications were indistinguishable from each other. The latter result suggests that the additional uncertainty in the model predictions of annual survival, mortality, harvest, and disturbance status contributes little to the overall uncertainty in the estimates of the annual BA means.

The third conclusion is that because 5 year ΔBA is usually small relative to BA 5 yr in the past and because the uncertainty in ΔBA predictions is small relative to the natural variation in BA among plots, ΔBA appears to be an appropriate quantity to use as the basis for updating.

The final conclusion, and the most important one with respect to the objective of the study, is that simple updating techniques are not only feasible, but also relatively easy to implement, and they produce acceptably unbiased and precise estimates of both annual BA means and standard errors of the means for large areas.

Although these updating techniques were adequate for their intended purpose, there may be opportunities to improve them. The IMPUTE technique might be enhanced as the result of investigations in three areas: selection of the most appropriate expanded geographical area from which to obtain imputation source data for small areas that include few plots; creation of similarity groupings based on additional variables; and selection of the best size of similarity groupings of plots, width of the predictor variable intervals, or both. The model-based updating techniques might be enhanced with more sophisticated models, calibration data obtained for shorter measurement intervals, and incorporation of additional predictor variables, particularly climatic variables.

In terms of quality of estimates versus ease of use, variations of the IMPUTE technique are worthy of serious consideration. Their only inherent disadvantages for this application are their dependence on 5 yr average annual growth as a surrogate for annual growth, their limited applicability beyond the purpose for which they are implemented, and possible bias due to nonrepresentative samples for small areas. However, under most conditions, these disadvantages will likely be minimal or can be relatively easily overcome.

Table 3. Median values for 250 simulations of Pine County, Minnesota, timberland estimates of annual mean BA (m^3/ha) and standard error of the mean (m^3/ha).

			IMPUTE							
	SAMPLE100		Statewide		County		PREDICT		REMOTE	
Year	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1	15.58	0.60	15.61	0.61	15.64	0.62	15.62	0.60	15.57	0.60
2	15.51	0.62	15.40	0.63	15.33	0.66	15.48	0.61	15.60	0.61
3	15.23	0.63	15.22	0.64	15.21	0.66	15.37	0.62	15.35	0.62
4	15.17	0.65	14.97	0.65	14.91	0.65	15.11	0.62	15.18	0.63
5	15.52	0.66	15.04	0.66	15.11	0.67	15.23	0.64	15.32	0.65
6	15.52	0.65	15.12	0.67	15.30	0.67	15.32	0.65	15.39	0.65
7	15.47	0.66	15.12	0.68	15.33	0.69	15.28	0.66	15.48	0.66
8	15.64	0.67	15.21	0.68	15.47	0.68	15.27	0.67	15.59	0.67
9	15.78	0.67	15.53	0.68	15.77	0.66	15.48	0.66	15.69	0.65
10	15.75	0.67	15.39	0.67	15.70	0.65	15.41	0.65	15.72	0.66

Although the best model-based results were slightly superior to the best imputation results, particularly for smaller areas, model development required a greater resource investment than did development of the imputation procedure. However, model development should not be considered a serious obstacle, because its only requirements are simple models, a simple predictor variable, and standard estimation techniques. In addition, model development will be greatly facilitated as calibration data from fixed radius FIA plots at 5 yr intervals become available. Finally, the small additional investment necessary to construct the models may be further justified if the models can be used for additional purposes, such as longer term future predictions.

In conclusion, these analyses demonstrate that for BA, simple, unbiased, and precise updating techniques can be easily developed and implemented. Although BA is only one inventory measurement of interest, its correlation with other variables such as volume and stems per unit area suggests wider applicability for these updating techniques. Thus, inventory estimation approaches based on updating plots measured in previous years should be considered viable and accurate alternatives.

Literature Cited

Befort, W.A. 2000. Change detection, stratification, and mapping for continuous forest inventory. P. 7–12 *in* 1999 Proc. of the Section on Statistics and the Environment, Am. Stat. Assoc., Alexandria, VA. 148 p.

Belcher, D.W., M.R. Holdaway, and G.J. Brand. 1982. A description of STEMS, the stand and tree evaluation and modeling system. USDA For. Serv. Gen. Tech. Rep. NC-279, 18p.

Brand, G.J., M.D. Nelson, D.G. Wendt, and K.N. Nimerfro. 2000. The hexagon panel system for selecting FIA plots for an annual inventory. P. 8–13 *in* Proc. of the first annual forest inventory and analysis symposium, McRoberts, R.E., et al. (eds.). USDA For. Serv. 57 p.

Conover, W.J. 1980. Practical nonparametric statistics, Ed. 2. Wiley, New York. 493 p.

Lessard, V.C., R.E. McRoberts, and M.R. Holdaway. (Submitted). Diameter growth models using Minnesota FIA data. For. Sci.

McRoberts, R.E. 1999. Joint annual forest inventory and monitoring system: the North Central perspective. J. For. 97(12):27–31.

McRoberts, R.E., and M.H. Hansen. 1999. Annual forest inventories for the North Central region of the United States. J. Agric. Biol. Environ. Stat. 4:361–371.

Miles, P.D., C.M. Chen, and E.C. Leatherberry. 1995. Minnesota forest statistics, 1990. Revised. USDA For. Serv. Res. Bull. NC-158. 138 p.

Myers, C.C., and T.W. Beers. 1971. Point sampling and plot sampling compared for forest inventory. Res. Bull. 877. Purdue Univ. Agric. Exp. Sta., Lafayette, IN. 81 p.

Reams, G.A., and J.M. McCollum. 2000. The use of multiple imputation in the southern annual forest inventory system. P. 228–233 *in* Integrated tools for natural resources inventories, Proc. of the IUFRO conf., Hansen, M.H., and T. Burk (eds.). USDA For. Serv. Gen. Tech. Rep. NC-212. 744 p.

Reams, G.A., and P.C. Van Deusen. 1999. The southern annual forest inventory system. J. Agric. Biol. Environ. Stat. 4:346–360.

Rubin, D.B. 1987. Multiple imputation in non-response surveys. Wiley, New York, 258 p.

Rubin, D.B., and N. Schenker. 1986. Multiple imputation for interval estimation for simple random samples with ignorable nonresponse. J. Am. Stat. Assoc. 81:366–374.

Sande, I.G. 1982. Imputation in surveys: coping with reality. Am. Stat. 36:145–152.

Sande, I.G. 1983. Hot-deck imputation procedures. P. 334–350 *in* Incomplete data in sample *surveys*, Madow, W.G., and I. Olkin, (eds.). Academic Press, New York.

Spencer, J.S., Jr. 1982. The fourth Minnesota forest inventory: timber volume and projections of timber supply. USDA For. Serv. Res. Bull. NC-57. 72 p.

USDA Forest Service. 1970. Operational procedures. For. Serv. Handb. 4809.11, Chapter 10:11.1-1–11.1-3. USDA Forest Service, Washington, DC.

Van Deusen, P.C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Can. J. For. Res. 27:379–384.

Van Deusen, P.C., T.R. Dell, and C.E. Thomas. 1986. Volume estimation from permanent horizontal points. For. Sci. 32(2):415–422.

White, D., J. Kimerling, and S.W. Overton. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartog. Geog. Info. Sys. 19:5–22.