The Southern Annual Forest Inventory System

Gregory A. REAMS and Paul C. VAN DEUSEN

The Southern Annual Forest Inventory System (SAFIS) is in various stages of implementation in 7 of the 13 southern states serviced by the Southern Research Station. The SAFIS design is an interpenetrating design where the n units (1/6 acre plots) are divided into $k = 5$ panels, each panel containing $m = n/k$ units. Panel 1 plots are measured in year 1, panel 2 in year 2, etc., such that all plots have been visited by the end of year 5. The panel cycle is repeated into perpetuity. Each panel, in effect, is a 5-year periodic survey with complete overlap of sample units. Numerous estimation schemes are possible, and we explore five possible options. The five options are (1) use existing periodic inventory programs to produce 5-year survey estimates by adjusting all five panels to a common year, (2) analyze each annual panel independently, (3) produce 5-year estimates by combining the five panel estimates by varying the weight given to each panel, (4) base inventory estimates on mixed estimation where actual and predicted values are combined, and (5) use imputation techniques such that unmeasured plots are filled in with imputed plots. A two-phase method for forest area estimation that uses the known map marginals from a thematic map is presented as an alternative to photo interpretation-based estimates.

Key Words: Annual survey; Area estimation; Imputation; Interpreting panel design; Time series modeling.

1. INTRODUCTION

The Forest Inventory and Analysis (FIA) unit of the Southern Research Station is responsible for providing forest inventory estimates for 13 southern states. Since the inception of the Southern FIA unit in 1933, the standard mode of operation has been to conduct periodic inventories by state on an approximate 10-year cycle. Prior to 1997, two states were inventoried concurrently and, on average, it required 2 years to complete the field work in each state. Thus, the data are accurate for several years but become increasingly unreliable thereafter.

The periodic inventory system results in survey data that are disjoint in time and space across the South. With ever increasing use of Southern forests, there is the need for reliable

Gregory A. Reams is the Head of Statistical Techniques and Remote Sensing, Forest Inventory and Monitoring, USDA Forest Service, Southern Research Station, P.O. Box 2680, 200 Weaver Boulevard, Asheville, NC 28802 (E-mail: greams/srs_fia@fs.fed.us). Paul C. Van Deusen is Principal Research Analyst with NCASI, Tufts University, Anderson Hall, Medford MA 02155.

©1999 American Statistical Association and the International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 4, Number 4, Pages 346–360

resource information on an annual basis across the region. In response to this need, the Southern Annual Forest Inventory System (SAFIS) was initiated in 1997. SAFIS represents a significant shift in the philosophy of forest inventories for the South.

Area estimation under the periodic inventory system has used a two-phase sample of aerial photo points and permanent ground plots. First, the area of forest in each county of a state is estimated by photo interpretation of a large number of photo points. There is usually one photo point per approximately 230 acres. Double sampling is used to estimate forest area, whereby the large number of photo points are interpreted and a smaller sample of ground plots are used as ground checks. The sample of ground plots (truths) are used to correct for photo interpreter misclassification and date of photography.

Measurements on an approximately 1/6-acre ground plot provide the basic mensuration data used in determining per acre estimates of forest attributes. Estimates of inventory by condition class (land use, forest type, stand origin, stand size, stand density, and ownership category) and tree species group are the major objectives of the survey. Equally important are estimates of change that result from processes such as growth, mortality, and removals by condition class and tree species. These estimates are all derived from remeasurement of permanent ground plots established in previous survey cycles.

The Southern periodic survey is best described as a pure overlapping panel design with a single panel. This panel is measured about every 10 years. The dates of the last periodic survey by state and the implementation of SAFIS are listed in Table 1. Several states (e.g., North Carolina, Tennessee) are in the process of closing out the final periodic survey for that state. Other states are in various stages of implementing the annual survey. For example, Virginia is measuring panel 3 and Georgia is measuring panel 2 (Table 1). All states will be operating under the annual system by the end of the year 2000.

Table 1. Number of Forested Plots and Inventory Dates for Southern States. For the periodic inventory, there was one panel; under the annual inventory, there are five panels.

State	*Number of forested plots*	*Last periodic inventory*	*Annual inventory implementation*	*Panel*
Alabama	3,944	1990	1997	3
Virginia	3,200	1992	1997	3
Georgia	4,820	1997	1998	2
Tennessee	2,361	1999	1999	1
Louisiana	2,415	1991	1999	1
South Carolina	2,520	1993	1998	2
Kentucky	2,460	1988	1999	1
Arkansas	3,209	1995	1999	1
North Carolina	3,860	1999	2000	1
Florida	3,240	1995	2000	1
Mississippi	3,190	1994	2000	1
Texas	3,840	1992	2000	1
Oklahoma	1,500	1993	2000	1
Puerto Rico and Virgin Islands	125	1990	2000	a

[a]Puerto Rico and Virgin Islands will be periodic surveys.

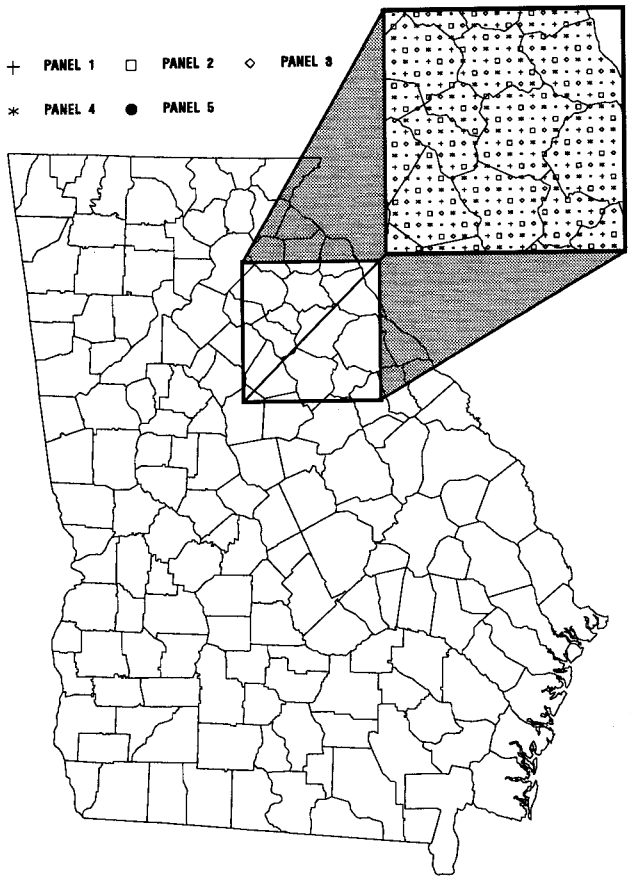

Figure 1. The Interpenetrating Panel Design of SAFIS in Georgia. All plots in panel 1 are measured in year 1, panel 2 plots in year 2, and so forth.

The Southern FIA unit uses a sampling intensity of approximately one ground plot per 5,760 acres. This sample intensity corresponds to a 3-mile systematic sample and, for analysis purposes, the sample is treated as a simple random sample (SRS) without replacement. It is generally accepted within the forest inventory community that SRS computations provide a good approximation, especially when the distribution of the variables in the population is effectively random. Most importantly, the forest inventory community wants the equal-area representation across a state that a systematic sample provides.

2. THE SAMPLE DESIGN OF SAFIS

The SAFIS sampling design is a systematic sample of five interpenetrating grids, also known as annual panels to Southern FIA (Fig. 1). This is a completely overlapping design with the same forest plots (elements) measured every 5 years. The SAFIS panel design provides quite a bit of flexibility. Although the basic design is to have five nonoverlapping panels, it might be sensible to reshuffle panel assignments after the first 5 years to allow for

a variety of growth intervals to be measured. This reshuffling could be restricted so that no plot is measured in less that 2 or 3 years since little change usually occurs over short time periods. Reshuffling would provide plot information at varying intervals of remeasurement and would provide an estimation mechanism to decrease the likelihood of year effects being portrayed as panel effects. Many model-based users of the survey will obviously advocate reshuffling while design-based advocates will not. The SAFIS program has not committed to reshuffling, and the remainder of our paper assumes no reshuffling.

The Southern FIA unit has historically used a completely overlapping single panel design for periodic inventories and is implementing a similar design with the new annual system. For definitional purposes, a panel denotes a sample in which the same elements (plots in this case) are measured on two or more occasions. Panel designs permit studies of individual change and therefore allow for the accounting of gross change that would be masked in a nonoverlapping design. In Southern's FIA program, the use of a panel design is largely due to the importance of estimating gross changes and trends in growth, mortality, and removals. As compared to results from two independent samples, the variance of the difference between means is reduced to the degree that the means from overlapping units are positively correlated (Van Deusen 1996a). With no overlap, the variance of the difference is $2S^2/n$, as compared to $(1 - R)(2S^2/n)$ for complete overlap (Kish 1965).

Usually in forest inventory, there is stability over time for many plot and tree characteristics. With pressure to balance timeliness, sampling errors, and budgets, a complete overlap of sample elements provides the greatest potential for variance reduction. For example, total cubic foot volume per acre is an inventory statistic of considerable importance. From the most recent survey in Georgia, the correlation between volume at times 1 and 2 is 0.498. Thus, with complete overlap, the variance of the difference is reduced by approximately half (i.e., $1 - 0.498$) over that of two independent samples.

A number of options exist for analyzing the SAFIS panel design. As has been the tradition in FIA, the program is focusing on the desired inventory-reporting format and significant ancillary uses of the data. Programmatic review has identified the following priorities. These priorities proceed from currently implementable to reasonably implementable in the short term to an applied research problem where implementation is several years away.

The top priority is to provide standard inventory estimates every 5 years. This suggests using SRS theory to compute estimates of population totals, means, and associated variances. This option describes Southern FIA's current system. In the typical southern state, it takes 2–3 years to collect and process the 10-year periodic survey. Under SAFIS, the period for sampling all plots in a state is extended to 5 years using the interpenetrating panels. The survey can then be processed as if it were a periodic survey, ignoring the fact that some plots (panels 1 and 2) are several years older than normal when compared to the traditional periodic survey. We'll call this the traditional periodic option, and it can rely on current software and estimation techniques that use simple random sampling (SRS). This option is favored by many in the traditional FIA community.

The second option to produce the standard FIA inventory report is to consider the fact that the SAFIS design is an interpenetrating design. Under this approach, each annual

panel can be analyzed independently and estimates can be produced on an annual basis. The five annual estimates can be combined by treating the design as a replicated sample (Mahalanobis 1946; Deming 1956; Jones 1956; Kish 1965). When all panels have the same number of member plots and equal weighting is used, then options 1 and 2 yield identical results.

Under options 1 and 2, another possible analysis alternative is to use only the most recent 3 years (panels) of SAFIS data for any analysis. This alternative is especially useful during the initial implementation of SAFIS because the survey program and its users will certainly expect inventory estimates based on the first three panels. Using less than five panels increases the variance of the estimates and will restrict some or perhaps all of the analyses to larger areas than users are accustomed to. However, the estimates will be based on reasonably current data. Incorporating older data (i.e., >3 years) should improve the estimates, and we discuss methods to do so under options 3, 4, and 5.

The development of SAFIS is related to the fact that many key users of FIA information have found the inventory estimates to be less reliable after 3 or more years. This is due to a variety of reasons, including natural disturbance such as mortality related to hurricanes, ice storms, or insect epidemics, as well as human-induced changes such as forest harvesting. This suggests that combining all five panels under options 1 and 2 may not be desirable for states with high disturbance rates. Another option (option 3) that can be easily implemented is to use a weighting scheme whereby the most recent panel estimates are given greater weight than the most dated ones.

An additional option (option 4) to consider in the analysis is that each panel is remeasured on a 5-year rotation. As shown earlier, there is serial correlation between e_1 and e_6, where 1 indicates current year estimates and 6 indicates 5 years prior. In this situation, time series models can be developed to incorporate time trend information into the survey estimates.

Finally, there is the need to consider the estimation of small domain means (Van Deusen 1996b). For each annual panel, the plot intensity is one ground plot per 28,800 acres. At this level of sampling intensity, survey estimates for a subregion within a state can be improved if reasonable values can be imputed for unmeasured plots (panels). There are many analyses by users of the Southern FIA data on a subregional (multicounty) basis. In this setting, it is desirable and necessary to improve precision by any means possible. Imputation (option 5) provides a cost-effective solution for improving precision for multicounty analysis of FIA data.

3. ESTIMATION TECHNIQUES

The Southern FIA unit currently analyzes the periodic survey by treating the data as if it were a simple random sample. Under option 1 as given in the previous section and assuming there are no panel effects, the SAFIS design can rely on SRS theory to provide population estimates across panels.

Consider Y_j to be the value of the y variable for the jth population element; in FIA's design, this is a ground plot of approximately 1/6 acre. The mean is

$$\bar{y} = \frac{y}{n} = \frac{1}{n}\sum_{j}^{n} y_j$$

and the variance is

$$\mathrm{var}(\bar{y}) = (1-f)\frac{s^2}{n},$$

where

$$s^2 = \frac{1}{n-1}\sum_{j}^{n}(y_j - \bar{y})^2$$

and $f = n/N$, the finite population correction, which is usually ignored given the large sample sizes associated with forest inventories. The standard error (SE) of $\bar{y}$ is the square root of its variance,

$$\mathrm{SE}(\bar{y}) = \sqrt{\mathrm{var}(\bar{y})} = \sqrt{1-f}\frac{s}{\sqrt{n}}.$$

Most users of FIA inventory data want to estimate population totals, and we provide totals of the Y_j variables. The estimator of Y is $N(\bar{y})$ and its standard error is estimated by

$$\mathrm{SE}(N\bar{y}) = N\,\mathrm{SE}(\bar{y}) = N\sqrt{1-f}\frac{s}{\sqrt{n}}.$$

In addition to estimating current inventory, a major goal for FIA is to estimate the difference between two survey periods. The Southern Research Station has used two completely overlapping samples for over four decades. The variance of the difference between time 2 and time 1 survey periods is

$$\mathrm{var}(x_2 - x_1) = \frac{1-f}{n}\left(s_{x_2}^2 + s_{x_1}^2 - 2s_{x_2x_1}\right)$$

or, if using $d_j = (x_2 - x_1)$,

$$s_d^2 = \frac{1}{n-1}\left(\sum_{j}^{n} d_j^2 - \frac{\left(\sum d_j\right)^2}{n}\right).$$

4. REPLICATED SAMPLING (INTERPENETRATING DESIGN)

Option 2 for analyzing the SAFIS design recognizes that each annual panel can be analyzed independently of all other panels. Compute the desired statistics x_γ for each of the c replications (panels). Each x_γ is an estimate of the corresponding population value. The mean of all panels is

$$\bar{x} = \frac{1}{c}\sum_{\gamma}^{c} x_\gamma \tag{4.1}$$

and the variance is

$$\mathrm{var}(\bar{x}) = \frac{1-f}{c(c-1)} \sum (x_\gamma - \bar{x})^2 = \frac{1-f}{c(c-1)} \left[\sum x_\gamma^2 - \frac{\left(\sum x_\gamma\right)^2}{c} \right]. \tag{4.2}$$

The only requirements for replicated sampling are that the c replications be independent and that the c replications use the same sampling design and sampling fractions (Kish 1965).

To estimate the population mean $\bar{Y}$ with replicated sampling, we express its estimate as $x_\gamma = \bar{y}_\gamma$ with each replication (panel) and then estimate the overall sample mean $\bar{y}$ and its variance as given in Equations (4.1) and (4.2). To estimate the population total Y, the replicated statistic can be its estimate $x_\gamma = cFy_\gamma$ (F is the inverse of the sampling fraction $1/f$) based on the sample total y_γ in each replication (Kish 1965). Equations (4.1) and (4.2) can serve any statistic based on c replicated statistics x_γ, each of which alone is an estimate of the target population value. Options 2 and 1 produce identical results when all panels have the same number of member plots.

5. WEIGHTED ESTIMATES

Option 3 assumes that users will want to use data from all five annual panels. It makes sense to emphasize the freshest data in the estimation of population statistics. The 4- and 5-year-old panels contain information that is more valuable than no data at all, but these data are clearly dated and less reflective of current values than the most recent panels. This suggests using a weighted sample mean,

$$\bar{y}_w = \sum_{p=1}^{5} W_p \bar{y}_p,$$

where $\Sigma W_p = 1$ and $\bar{y}_w$ is equal to the sum of the p panel sample means $\bar{y}_p$. The sample mean is obtained separately and independently for each panel, and it is then multiplied by the weight of the panel. These products are summed over the p panels to obtain the weighted sample mean. The variance of this weighted mean is obtained by combining the separate variances of the panel means. The variance of each panel mean is multiplied by the square of the panel weight and the products are added over the p panels, i.e.,

$$\mathrm{var}(\bar{y}_w) = \sum W_p^2 \mathrm{var}(\bar{y}_p).$$

Weights are frequently chosen to represent the proportions of the population elements in each panel and $W_h = N_h/N$. The weights in the annual inventory are chosen to represent the age of the panel, with the greatest weight given to the most recent panels and the least weight given to the oldest.

Several obvious choices of weights are 0.4, 0.3, 0.2, 0.1, 0; 0.4, 0.2, 0.2, 0.1, 0.1; 0.3, 0.2, 0.2, 0.2, 0.1; 0.3, 0.3, 0.2, 0.1, 0.1. The long-standing periodic forest inventory system within FIA most often collects current cycle data over a 2- or 3-year period per state. Equal weight is applied to all plots in estimating population statistics and, if this precedent were

followed, would suggest a weighting scheme of 0.2, 0.2, 0.2, 0.2, 0.2, i.e., equal weighting. This is not as logical as giving current information greater weight than older information.

The population total $N\bar{Y}_w$, where N is a constant, estimated independently of the sample results can be estimated by $N\bar{y}_w$, and the variance and standard error of this estimate are

$$\mathrm{var}(N\bar{y}_w) = N^2\mathrm{var}(\bar{y}_w),$$
$$\mathrm{SE}(N\bar{y}_w) = N\,\mathrm{SE}(\bar{y}_w).$$

6. TIME SERIES AND MIXED ESTIMATION

All prior options assume that no serial correlation exists between observations. Option 4 recognizes that this is not a true reflection of our sample because all panels will be remeasured every 5 years. There is also the desire to combine modeled plot measurements with actual measurements. There are a number of time series models and assumptions that can be used to improve the efficiency of our estimates. In particular, the Kalman filter (Kalman 1960) and the mixed estimator (Theil 1971) have been suggested (Van Deusen 1996a) for this particular application.

For mixed estimation, Van Deusen (1996a) suggested the following model is justified for a systematic sampling analysis assuming that the population is in random order (Cochran 1977):

$$\bar{y}_t = \mu_t + \bar{e}_t, \qquad t = 1, \ldots, T, \tag{6.1}$$

where $\bar{y}_t$ is the observed sample mean, μ_t is the true population mean, and $\bar{e}_t$ is the error term with mean zero and variance σ_t^2/n_t. Model (6.1) assumes no serial correlation between $\bar{e}_t$ and $\bar{e}_{t-j}$ for all j. With SAFIS, the same subset of plots will be measured after 5 years; therefore, it is difficult to justify the assumption of no serial correlation (Van Deusen 1996a). The above analysis could be modified to incorporate serial correlation to improve the efficiency of the estimate. The estimators would be similar to those for sampling with partial replacement (Ware and Cunia 1962; Scott and Kohl 1994; Van Deusen 1989).

To incorporate growth projections from models and the time series nature of the data, consider the model

$$x_t - x_{t-1} = \mu_t - \mu_{t-1} + v_t, \qquad t = 2, \ldots, T, \tag{6.2}$$

where v_t is an error term with mean zero and variance λ_t^2/m_t, where m_t denotes the sample size for the x's. The x's are restricted to values that are modeled or in some way estimated on the unmeasured plots for the current year. Equation (6.2) provides an independent estimate of change in the population mean over time based on the modeled x's. There are advantages to the approach of modeling change in that any additive bias disappears.

The mixed estimator (Theil 1971) provides a method for combining the estimated or modeled information [Eq. (6.2)] and the measured plot data [Eq. (6.1)]. Van Deusen (1996a) gives a full accounting of the mixed estimator model specification. The mixed estimator

incorporates prior information into the estimation process to link parameter estimates for adjacent time periods. With the mixed estimator, the parameters are treated as random rather than fixed, as would be the case with the Kalman filter. However, the mixed estimator is more easily implemented because starting values for time 0 are not required as with the Kalman filter. The initialization period with the Kalman filter can be problematic for forest inventory use given the short time periods for reporting in SAFIS.

7. IMPUTATION

Imputation methods are a fifth estimation option. Imputation is seemingly new to forest inventory since there are few publications that formally address the topic. However, upon further inspection, it is clear that the profession has practiced imputation for several decades, most notably in Scandinavia (Poso 1978; Holm, Hagglund, and Martensson 1979). In the United States, many inventory systems have used imputation, although under the label of modeling rather than imputation. The data that are modeled or imputed are treated as actual, and the business of producing inventory estimates proceeds as normal.

Historically, imputations have occurred via different implementation schemes. This is true whether whole plots are imputed using matched plot procedures as used and internally documented by the Southeastern Research Station (Cost, personal communication) or if plot volumes are predicted using inventory projection models similar to the Stand and Tree Evaluation and Modeling System (STEMS) (Belcher, Holdaway, and Brand 1982) as documented by the North Central Research Station (Leatherberry, Spencer, Schmidt, and Carroll 1995). The examples just listed fall under the definition of single imputation, i.e., filling in a value for each missing value. The use of models within STEMS to update or estimate growth and volume on undisturbed plots is an example of mean imputation as defined by Rubin (1987). Thus, imputation has been used for several decades for inventory estimation without it being called such. Obviously, the method works well since inventory estimates based on imputed or projected plots are part of standard operating procedures within FIA (Leatherberry et al. 1995; Cost, personal communication).

Prior applications of imputation include what are popularly known in the Southern Research Station as fly plots (Cost, personal communication). Fly plots are ground samples in inaccessible areas such as roadless areas in the coastal swamps and wetlands. The procedure for matching donor plots to the fly plots is conceptually similar to the Census Bureau's hot-deck procedure (Sande 1983). The matching of a donor to a recipient plot is based on forest type, stand size, stand age, disturbance, and physiographic class. Ideally, these matches are done without replacement to minimize the usage of any donor plot.

Because both the model- and matching-based implementations are using a form of single imputation, the variance estimates are biased on the low side. The situation is likely more pronounced when using mean imputation because there is no variability in the predicted (imputed) value given the same set of predictants (Ek, Robinson, Radtke, and Walters 1997). With plot matching procedures, the variability of imputed values is greater given there are multiple donors that meet the match criteria (Rubin 1987).

One way to improve estimates based on regression model imputations is to include the addition of a residual value chosen by hot deck or some other random process. In this case, the imputation is

$$Y_{\text{imp}} = \hat{Y} + \hat{e},$$

where $\hat{Y}$ is the predicted value obtained from the fitted model, which is based on the complete observations, and $\hat{e}$ is the estimated residual, which may be obtained by hot deck from the actual residuals of the fitted values or randomly generated using the estimated distribution of the residuals (Sande 1982).

Inventory estimates, especially small domain estimates, can likely be improved by updating plots through models or matching plot procedures. In this way, the sample size can be increased by using out-year plots, thereby improving both small and large domain estimates. There are two obvious choices for the implementation of imputation in the Southern FIA program. One is to use models for updating variables; the second is to use hot-deck (matching) procedures to fill in missing plots. Either modeling or matching procedures will result in having current data for one panel and imputed data for the other four panels.

Multiple imputation retains the virtues of single imputation and corrects its major flaws. Rather than impute a single value, the concept is to impute several values, say m. As Rubin (1987) states, there are several advantages to multiple imputation over single imputation. The two advantages that stand out in forest inventory applications are, first, when imputations are randomly drawn in an attempt to represent the distribution of the data, multiple imputation increases the efficiency of estimation. Second, when the multiple imputations represent repeated random draws under a model for nonresponse, valid inferences can be obtained by combining complete-data inferences in a straightforward manner. Because multiple imputation maintains the diversity that is inherent to the data, inventory users and specialists can reach valid inferences using only familiar complete-data tools (Van Deusen 1996b). The basic framework for multiple imputation is covered in the following section.

8. ESTIMATION WITH MULTIPLE IMPUTATION

Let Q be the quantity of interest in the survey. It could, e.g., be $\bar{y}$ or some other parameter that is of interest. Q is a k-dimensional row vector, and we assume inferences for Q are based on the assumption that

$$(Q - \hat{Q}) \sim \mathrm{N}(0, U),$$

where $\hat{Q}$ is an estimate of Q. After generating m simulated completed datasets and analyzing each of them as if they were genuine complete datasets, we now have m estimates for Q and U, i.e., $\hat{Q}_{*1}, \ldots, \hat{Q}_{*m}$ and $\hat{U}_{*1}, \ldots, \hat{U}_{*m}$.

The m repeated complete-data estimates and associated complete-data variances for Q is

$$\bar{Q}_m = \sum_{i=1}^{m} \hat{Q}_{*i}/m,$$

which is the mean of means. The total variance of $\bar{Q}_m$ is estimated by

$$T_m = \bar{U}_m + (1 + m^{-1})B_m,$$

where

$$\bar{U}_m = \sum_{i=1}^{m} \hat{U}_{*i}/m$$

is the average of the m complete-data variances and

$$B_m = \sum_{i=1}^{m} (\hat{Q}_{*i} - \bar{Q}_m)(\hat{Q}_{*i} - \bar{Q}_m)/(m-1)$$

is the variance among the m complete-data estimates. Interval estimates when Q is a scalar are based on $\bar{Q}_m \pm kT_m^{1/2}$, where k is the appropriate percentile of the t distribution with degrees of freedom

$$v = \left[1 + \left(\frac{m}{m+1}\right)\frac{\bar{U}_m}{B_m}\right]^2 (m-1).$$

The Census Bureau's hot deck procedure provides a method for imputing values for unmeasured plots. The hot deck procedure uses categorical X to match donors to recipients, and the donors then give their values to the recipients. If more than m respondents are matches, then a subsample of m respondents can be drawn without replacement. If less than m matches are found, then one or more of the X-variables are made coarser. The choice of matching fields in both sequential and random choice procedures must be made considering likely sources of variation and the number of complete or eligible records available as potential donors. If too many fields are used for matching, the number of potential donors may be too small; if too few fields are used for matching, there is the risk of a poor match in the imputed records (Sande 1982). Following prior methods of plot imputation in the Southern forest survey, the candidate variables for categorical X are survey unit, forest type, stand size, stand age, and physiographic class.

With hot deck methods, the variance of the estimates in simple cases is known to be larger than the variance of the usual expansion estimates of means and totals. However, there may be a reduction in bias. Compared to some other methods of imputation, such as the use of models, hot deck methods should produce imputed data sets that appear more realistic and do a better job of reflecting distributional properties (Sande 1982).

Imputation using normal linear regression could be used for some important continuous variables. This option will be considered when more resources are available for implementation.

9. FOREST AREA ESTIMATION WITH SATELLITE DATA

Historically, the Southern FIA program has produced area estimates of forest type using a variation of double sampling. The process consists of interpreting a large number of sample points on aerial photographs and subsampling a proportion of the points on the ground. The

aerial photo sampling is referred to as phase I and is used to estimate the percent of the total area occurring in forest and nonforest subpopulation. The phase II samples are the FIA ground plots that provide the basic mensurational data used in determining individual tree and per unit area estimates of volume, increment, and yield.

SAFIS places greater demand for the rapid generation of land-use and land-cover maps for the southern United States. The current photo system for phase I area estimation has changed little over the last four decades and provides area estimates within the precision requirements of the FIA program. The photo method does have two shortcomings for the SAFIS program. First, although FIA can provide estimates of forest land down to the county level, the program cannot produce maps of forest and nonforest area. Second, it takes a considerable amount of time (up to a year per state) to photo interpret an entire state. A stated goal of SAFIS and the national FIA program is to eventually replace photo interpretation with digital satellite classification to address the two shortcomings of the photo system. An area estimation method that replaces photo interpretation with digital satellite data follows.

Satellite-based imagery provides the basic data needed to classify cover types of large areas in an automated, cost-effective, and timely manner. The three most commonly used satellite sensors are the thematic mapper (TM), the French Systeme Probatoire pour l'Observation de la Terre (SPOT), and Advanced Very High Resolution Radiometer (AVHRR). Presently, we concentrate on the use of TM classifications because TM has greater spectral resolution relative to SPOT and better spectral and spatial resolution compared to AVHRR.

To estimate map class area totals and variances, we use two-phase or double sampling, where the less accurate data is the map whose accuracy is in question and the more accurate but costly data is the FIA ground sample. The less accurate data is complete in that each map pixel has been classified.

A sampling scheme designed to evaluate and correct for map area misclassification is as follows: A sample of n points/pixels is located on the map and the true and map categories are determined for each point. The n points are allocated as a simple random sample. This results in a two-way contingency table where n_{ij} is the number of points in the sample whose true category is i and whose map category is j.

There is an important difference between using satellite-derived maps and aerial photos for this process. The satellite-derived thematic map allows us to know the actual map marginal probabilities, which can then be used as additional constraints in a maximum likelihood estimation process (Card 1982; Van Deusen 1994, 1996c). This reduces the variance of estimates of true map category proportions. Formulas for estimating the true probabilities of interest are given in Card (1982), along with variance estimates. Methods for estimating change in category proportions between two times are given in Van Deusen (1994), along with variance formulas. Coincidentally, the estimators for the true map proportions are the same for SRS or stratified sampling of map pixels. However, variance estimators are different under the two sampling strategies.

The thematic map accuracy assessment could be done with either stratified or SRS. Stratified sampling allows for taking different sample sizes within each map category and might be advantageous. Some categories are more important than others, e.g., forest is more important than nonforest for this survey. Likewise, less common forest types might require larger sample sizes to obtain the desired precision on estimates of true proportions. Stratified sampling would require that truth be obtained from high-resolution photography. Simple random sampling would be most appropriate when ground plots are used to determine truth. In this case, samples have no prior relationship with the thematic map strata and stratified sampling estimators for variance would be inappropriate (Czaplewski and Catts 1992). However, it is still important to use the known map marginals to reduce variance of the final estimates. When the reference (ground truth) plots are a simple random sample of the classified satellite image, a sample size of 500–1,000 plots is recommended (Czaplewski and Catts 1992).

10. CONCLUSIONS

For those familiar with the longstanding 10-year periodic FIA survey, the interpenetrating panel design of SAFIS is analogous to taking the large periodic survey and dividing it into five repeated smaller samples. The chief advantage of the annually repeated survey over the traditional periodic sample is that the separate annual samples provide information about variations that occur between the periods. This results in the ability to estimate annual and secular trends. Also, the sum of repeated surveys over the entire period can lead to better statistical inference than a single, concentrated, one-time survey.

Probability selection of time segments from an entire interval permits statistical inference from the sample to an average condition over the interval. In comparison, the traditional FIA periodic survey for an entire interval demands judgment and assumptions about the nature of variation over the entire interval. The one-shot survey is exposed to more risks of secular and catastrophic variations, known or unknown. The strength of the SAFIS repeated survey is that it relies on averaging out the variations over the repeated survey (Kish 1965).

The SAFIS design provides for increased flexibility of measurement and estimation. A number of estimation techniques have been presented, with each estimation option having a particular set of goals and assumptions. Each option presented provides a different process for averaging each of the annual panels over the repeated survey cycle. For example, estimates of inventory based on pooling all data to a common year or averaging estimates over all five panels leads to identical estimates, given the same number of member plots for the analyses. For users that wish to rely on the most current panels and discount the oldest data (panels), inventory estimates based on a weighted average have been presented. The use of the interpenetrating annual panels will encourage the use of models to estimate unmeasured plots (out-year panels). The mixed estimator illustrates the time series nature of the SAFIS design and provides a model for combining predicted and measured values. Many users of FIA data are interested in small domain estimates. Given the paucity of data, especially for small domains for any year, imputed or predicted values of unmeasured plots can be

used to increase the amount of information available. For this purpose, conceptually simple imputation techniques for both small and large domain estimation have been highlighted.

With the increased turnaround time of SAFIS, development of automated procedures for forest typing (mapping) and area determination are needed. Wall-to-wall forest typing and area determination from remotely sensed digital satellite data provide a viable solution. Estimates of forest area using the known map marginals from thematic maps is discussed.

[Accepted May 1999.]

REFERENCES

Belcher, D. W., Holdaway, M. R., and Brand, G. J. (1982), "A Description of STEMS, the Stand and Tree Evaluation and Modeling System," General Technical Report NC-79, U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Card, D. H. (1982), "Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy," *Photogrammetric Engineering and Remote Sensing,* 48(3), 431–439.

Cochran, W. G. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

Czaplewski, R. L., and Catts, G. P. (1992), "Calibration of Remotely Sensed Proportion or Area Estimates for Misclassification Error," *Remote Sensing of the Environment,* 39, 29–43.

Deming, W. E. (1956), "On Simplifications of Sampling Design Through Replication With Equal Probabilities and Without Stages," *Journal of the American Statistical Association,* 51, 24–53.

Ek, A. R., Robinson, A. P., Radtke, P. J., and Walters, D. K. (1997), "Development and Testing of Regeneration Imputation Models for Forests in Minnesota," *Forest Ecology and Management,* 94, 129–140.

Holm, S., Hagglund, B., and Martensson, A. (1979), "A Method for Generalization of Sample Tree Data From the Swedish National Forest Survey," Report 25, Swedish University of Agricultural Sciences, Department of Forest Survey, Umea, Sweden.

Jones, H. L. (1956), "Investigating the Properties of a Sample Mean by Employing Random Subsample Means," *Journal of the American Statistical Association,* 51, 54–83.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering,* 82, 34–45.

Kish, L. (1965), *Survey Sampling,* New York: Wiley.

Leatherberry, E. C., Spencer, Jr., J. S., Schmidt, T. L., and Carroll, M. R. (1995), "An Analysis of Minnesota's Fifth Forest Resources Inventory, 1990," Resource Bulletin NC-165, U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Mahalanobis, P. C. (1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," *Journal of the Royal Statistical Society,* 109, 325–370.

Poso, S. (1978), "National Forest Inventory in Northern Finland," in *National Forest Inventory,* Proceedings of IUFRO Subject Group S4.02 Meeting, Bucharest, Romania.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys,* New York: Wiley.

Sande, I. G. (1982), "Imputation in Surveys: Coping With Reality," *The American Statistician,* 36(3), 145–152.

——— (1983), "Hot-Deck Imputation Procedures," in *Incomplete Data in Sample Surveys,* eds. W. G. Madow and I. Olkin, New York: Academic Press, pp. 334–350.

Scott, C. T., and Kohl, M. (1994), "Sampling With Partial Replacement and Stratification," *Forest Science,* 40, 30–46.

Theil, H. (1971), *Principles of Econometrics,* New York: Wiley.

Van Deusen, P. C. (1989), "Multiple-Occasion Partial Replacement Sampling for Growth Components," *Forest Science,* 35, 388–400.

——— (1994), "Correcting Bias in Change Estimates From Thematic Maps," *Remote Sensing of the Environment,* 50, 67–73.

——— (1996a), "Incorporating Predictions Into an Annual Forest Inventory," *Canadian Journal of Forest Research,* 26, 1709–1713.

——— (1996b), "Annual Forest Inventory Statistical Concepts With Emphasis on Multiple Imputation," *Canadian Journal of Forest Research,* 27, 379–384.

——— (1996c), "Unbiased Estimates of Class Proportions From Thematic Maps," *Photogrammetric Engineering and Remote Sensing,* 62(4), 409–412.

Ware, K. D., and Cunia, T. (1962), "Continuous Forest Inventory With Partial Replacement of Samples," *Forest Science,* Monograph 3.