SAFIS AREA ESTIMATION TECHNIQUES

Gregory A. Reams

ABSTRACT.—The Southern Annual Forest Inventory System (SAFIS) is in various stages of implementation in 8 of the 13 southern states served by the Southern Research Station of the USDA Forest Service. Compared to periodic inventories, SAFIS requires more rapid generation of land use and land cover maps. The current photo system for phase one area estimation has changed little over the last four decades and provides area estimates within the precision requirements of the FIA program. A stated goal of the national FIA program is to eventually replace photo interpretation with digital satellite classification because the photo system cannot produce maps of forest and nonforest area, and it takes an enormous amount of time to photo interpret the phase one photo plots. Using automated classification procedures for TM satellite data, we anticipate that the time to complete phase one will decline and wall-to-wall maps will be available. In the interim period of switching to satellite data, the photo system must be modified to provide current estimates of inventory. A method being used by Southern FIA is documented.

INTRODUCTION

Historically, the Southern FIA program has produced forest area estimates using a variation of double sampling. The process consists of interpreting a large number of sample plots on aerial photographs and subsampling a proportion of the plots on the ground. The aerial photo sampling is referred to as phase one and is used to estimate the percent of the total area occurring in forest and nonforest subpopulations. The phase two samples are the FIA ground plots that provide the basic mensurational data used to further stratify forest area and estimate timber volume, growth, mortality, and removals.

The Southern Annual Forest Inventory System (SAFIS) requires rapid generation of land use and land cover maps for the southern United States. The current photo system for phase one 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 annual inventory program. First, although the photo method can provide estimates of forestland down to the county level, the method cannot produce maps of forest and nonforest area and distribution. 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. In the interim, the following photo-based procedures are being used to estimate forest area under the annual inventory system in the Southern FIA program.

USING DOUBLE SAMPLING FOR AREA ESTIMATION

Currently the Southern FIA is using a double sample to estimate forest area. Frayer and Furnival (1999) provide a chronology of how double sampling for stratification has been implemented nationwide by FIA. Cochran (1977) and De Vries (1986) provide statistical references for estimation when employing a double sample. Estimates of timberland area are based on forest and nonforest interpretation of a large number of plots on aerial photos and a smaller sample of ground plots. There is approximately 1 photo plot per 230 acres across the South. The current definition of forest in FIA is land 1 acre in size, 120 feet wide, and at least 10 percent stocked by forest trees of any size, or formerly having had such tree cover and not currently developed for nonforest use.

The photo interpretation points are arranged in a 5 x 5 grid where one of the photo plots is spatially coincident with a phase two ground plot. The current phase two ground plots are conceptually distributed on a 3 x 3 mile grid. In addition to the double sample that occurs with each FIA ground plot, there are intensification plots arranged on a 3 x 6 mile grid. Intensification plots are used to increase sample size for forest and nonforest ground truth type calls. An intensification plot is simply a spatially coincident photo plot and field truth location. This results in 173.6 ground plots and 86.8 intensification plots (used to correct area estimates) per million acres. Combined, there are 260.4 (173.6 + 86.8) ground reference samples per million acres and 4,347.8 photo interpretation points per million acres. Thus, an approximate 6 percent field sample of the photo plots is used to form the double sample estimate of forest area.

A quick review of how forest area is determined using a periodic survey will naturally lead into modifications necessary to operate under the new annual panel system. For detailed reference information on the new annual panel system, see Reams and Van Deusen (1999) and Roesch and Reams (1999). To estimate forest area, several types of information are needed. First, the total area estimate and census water estimate by county are obtained from the U.S. Census Bureau. Census land is computed as census total area minus census water. Forest area ($\hat{A}_f$) is then computed as:

$$\hat{A}_f = \hat{P}'_f \times A_t \quad (1)$$

where

$$\hat{P}'_f = (\hat{P}_f \; x \; C_f) + (\hat{P}_n \; x \; C_n)$$

and, A_t is census land area,

$\hat{P}_f$ is the proportion of phase one photo plots in forest, $\hat{P}_n$ is the proportion of phase one photo plots in nonforest, and

$$C_f = \frac{\textit{number of plots correctly photo interpreted forest}}{\textit{total number of plots photo interpreted forest}}$$

and,

$$C_n = \frac{\textit{no. of plots photo interpreted as nonforest but are forest}}{\textit{total number of plots photo interpreted as nonforest}}$$

Assuming the following confusion matrix (table 1),

Table 1.—*Two-way contingency table where the diagonal elements represent correct classifications as compared to ground truth, and the off-diagonal elements represent plots with incorrect photo classification*

Grounds plots	Photo interpreted forest	Photo interpreted nonforest	Total
Forest	108	2	110
Nonforest	3	81	84
Total	111	83	194

the proportion of forest area ($\hat{P}'_f$) is then estimated as,

$$\hat{P}'_f = \frac{(\textit{no. points forest} \times C_f) + (\textit{no. points nonforest} \times C_n)}{\textit{total no. points photo interpreted}}$$

Assuming that 1,962 photo points were interpreted as forest and 1,288 photo points interpreted as nonforest for a total of 3,250 interpreted points, the proportion of forest area would be estimated as,

$$\hat{P}'_f = \frac{(1962 \times .973) + (1288 \times .024)}{3250} = .5969$$

Proportion of nonforest is simply 1- $\hat{P}'_f$.

Forest area (equation 1) is thus equal to 0.5969 x A_t. The variance of forest area is determined as follows,

$$\sigma^2(\hat{P}'_f) = \frac{(\hat{P}_f)(\hat{P}_n)}{n}(C_f - C_n)^2 + \left[\frac{(\hat{P}_f)^2 (C_f)(1-C_f)}{m_1}\right] + \left[\frac{(\hat{P}_n)^2 (C_n)(1-C_n)}{m_2}\right]$$

where, $\hat{P}_f$ = proportion forest, $\hat{P}_n$ = proportion nonforest, C_f = proportion correctly classified forest, C_n = proportion incorrectly classified nonforest, m_1 = number of forest field locations, m_2 = number of nonforest field locations, and n = number of photo points interpreted. Inserting values for each variable results in,

$$\sigma^2(\hat{P}_f') = \frac{(.6037)(.3963)}{3250}(.973 - .024)^2 + \left[\frac{(.6037)^2(.973)(1-.973)}{110}\right] + \left[\frac{(.3963)^2(.0241)(1-.0241)}{84}\right]$$

Completing the above mathematical operations, $\sigma^2(\hat{P}_f') = 0.00019729$ and the s.e. of the estimate is 0.014.

MODIFICATION OF CORRECTION FACTORS FOR ANNUAL SYSTEM

Because the Southern FIA program is dependent on the National Aerial Photography Program (NAPP) for acquisition of 1:40,000 scale photography that is flown on an approximate 5-year cycle, new photography will not be available on an annual basis. For the Southern annual survey, only one-fifth of the present number of phase two ground plots are remeasured in any one year. If relying solely on the FIA ground plots for the double sampling estimate of forest area, this results in approximately 35 ground truths per million acres. The Southern FIA program has additional ground truths arranged on a 3 x 6 mile grid, and if we assume one-fifth of these plots are visited every year, this equates to approximately 17 ground truths per million acres. Thus, in any given year, 52 ground truths per million acres are available for forest area estimation.

Because new photography is available on about a 5-year cycle, all photo plots will be interpreted as available, and will be completed in one calendar year. There will be 52 ground truths per million acres available per year, and this will be too small a sample size for reliable estimation at the county level because the average size of counties in most states in the South ranges from 200,000 to 350,000 acres. The reason for stating that the sample size for a panel (year) is too small for reliable estimation at the county level, is because the off-diagonal elements of the misclassification matrix will often be zero.

One remedy for this situation is to estimate forest area at the multi-county level based on 100 percent measurement of phase one photo interpretations but only 20 percent of the phase two ground plots (truths). Counties must be aggregated such that the misclassification matrix for the multi-county area has at least 200 ground truths per year. This number of ground truths usually ensures that the off-diagonals are not zero.

If one were to develop a correction factor on one panel (20 percent of the ground plots) of data, a reasonable recommendation to group counties either by immediate adjacency, percent forest, or percent change in forest area. Because misclassification errors are more often related to percent forest area, this option is being developed and is favored by Southern Station analysts. Using western (survey unit 1) Tennessee as an example, aggregation of counties totaling at least 2.5 million acres is recommended (fig. 1). The resulting misclassification matrix and correction factors are then applied

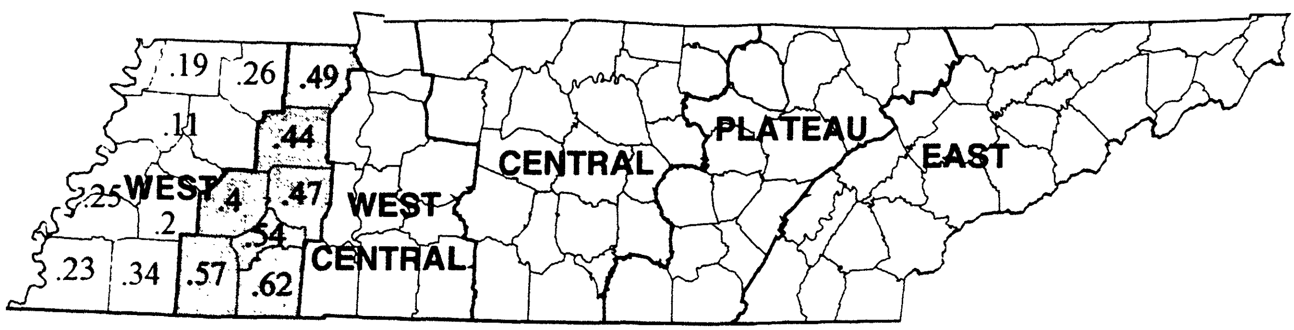

Figure 1.—*Example of how to group counties in survey unit 1, Tennessee for development of correction factors based on one panel of data. Shaded counties are aggregated for a common correction factor.*

individually to each county. Thus, a set of counties will have the same set of correction factors applied to each county's unique estimate of percent forest area from the phase one sample. In a recent application, counties were aggregated by survey unit and percent forest area class with class 1 (0-35 percent), class 2 (36-55 percent), class 3 (56-65 percent), class 4 (66-75 percent), and class 5 (76-100 percent). Forest area by county is determined then as (percent forest land x census land) as given previously in equation (1).

If estimating area based on multiple panels (years) of ground truths, there is the need to either assume that previous panel ground truths are correct, or revisit those plots or a subset of the previous panel of plots, since the forest/nonforest classification may have changed. Because of cost factors, the assumption made is that previous year panels or a subset of the previous year panel plots will not be visited. Assuming the same amount of field effort every year, the number of counties or the land area that must be aggregated each year for calculation of correction factors is cut in half in year two, a third in year three, and so on. By the time all five panels have been measured, we are back at the intensity of ground truths analysts are accustomed to under the periodic system. As many analysts well know, even with 100 percent (all five panels) measurement of ground plots, small counties are often combined because of small sample sizes. The need to combine counties for correction factors is thus a one-time adjustment with implementation of the five panel system.

USING SATELLITE DATA FOR AREA ESTIMATION

As stated earlier, the 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.

The three most commonly used satellite sensors are the Landsat Thematic Mapper (TM), the French Systeme Probatoire pour l'Observation de la Terre (SPOT), and Advanced Very High Resolution Radiometer (AVHRR). We have concentrated on the use of TM classifications because TM has greater spectral resolution relative to SPOT and better spectral and spatial resolution compared to AVHRR (Wynne *et al.* 2000).

To estimate map class area totals and variances, we use two-phase or double sampling, where the less accurate data are the map whose accuracy is in question and the more accurate but costly data are the FIA ground plot. The less accurate data are 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 be used as additional constraints in a maximum likelihood estimation process (Card 1982, Van Deusen 1996, Reams and Van Deusen 1999). 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 and variances between two times are given in Van Deusen (1994). The estimators for the true map proportions are the same for simple random sampling or stratified sampling of map pixels. However, variance estimates are different under the two sampling strategies.

In a pilot study in central Georgia, a winter cloud-free TM scene was classified and statistical estimates of forest and nonforest area were derived using the methods of Card (1982). Estimates of percent forest and nonforest from the classified TM scene compare quite favorably to FIA survey cycle six estimates of forest area based on the photo method. The FIA cycle six survey for central Georgia indicates that 68.7 percent of survey unit 3 is in forest (Thompson 1989), and the estimate based on TM data indicates that 69.4 percent of the scene is forest (Reams and McCollum, in review). Variance estimates are comparable to those derived from the photo-based method that uses standard double sampling for stratification statistical estimates. With the wall-to-wall TM classification, it is important to use the known map marginals to reduce variance of the final estimates (Card 1982, Reams and Van Deusen 1999).

LITERATURE CITED

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**. New York: Wiley. 413 p.

De Vries, P.G. 1986. **Sampling theory for forest inventory**, New York: Springer-Verlag. 399 p.

Frayer, W.E.; Furnival, G.M. 1999. **Forest survey sampling designs: a history**. Journal of Forestry. 97(12): 4-10.

Reams, G.A.; Van Deusen, P.C. 1999. **The southern annual forest inventory system**. Journal of Agricultural, Biological, and Environmental Statistics. 4(4): 346-360.

Reams, G.A.; McCollum, J.M. (In review). **Estimates of forest area using thematic map data**. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Roesch, F.A.; Reams, G.A. 1999. **Analytical alternatives for an annual inventory system**. Journal of Forestry. 97(12): 33-37.

Thompson, M.T. 1989. **Forest statistics for Central Georgia, 1989**. Resour. Bull. SE-105. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 51 p.

Van Deusen, P.C. 1994. **Correcting bias in change estimates from thematic maps**. Remote Sensing of the Environment. 50: 67-73.

Van Deusen, P.C. 1996. **Unbiased estimates of class proportions from thematic maps**. Photogrammetric Engineering and Remote Sensing. 62(4): 409-412.

Wynne, R.H.; Oderwald, R.G.; Reams, G.A.; Scrivani, J.A. 2000. **Optical remote sensing for forest area estimation**. Journal of Forestry. 98(5): 31-36.

ABOUT THE AUTHOR

Gregory A. Reams is a Supervisory Mathematical Statistician, Southern Research Station, Forest Inventory and Analysis Program, USDA Forest Service.

The author thanks Bill Bechtold, Ray Sheffield, and Mike Thompson for review and suggestions that improved the manuscript.