PEER-REVIEWED ARTICLE

PRACTICAL PAPER

A Comparison of Various Estimators for Updating Forest Area Coverage Using AVHRR and Forest Inventory Data

Francis A. Roesch, Jr., Paul C. Van Deusen, and Zhiliang Zhu

Abstract
Various methods of adjusting low-cost and possibly biased estimates of percent forest coverage from AVHRR data with a subsample of higher-cost estimates from the USDA Forest Service's Forest Inventory and Analysis plots were investigated. Two ratio and two regression estimators were evaluated. Previous work (Zhu and Teuber, 1991) finding that the estimates from the two different data sources differed the most in highly fragmented land-use areas led to an investigation into improving the estimates through the use of independently derived estimates of population density. It is concluded that reasonable updates of percent forest area could be obtained through both the ratio and regression estimators and that use of population density as a surrogate for land-use fragmentation could help improve the estimators, although more direct methods of measuring land-use fragmentation might provide further improvement.

Introduction
The USDA Forest Service Southern Forest Experiment Station Forest Inventory and Analysis Unit (SOFIA) is responsible for surveying the forest resources in seven mid-south states and Puerto Rico. This responsibility was mandated by the Forest and Rangeland Renewable Resources Planning Act of 1974. Due to the extent of the assignment, surveys for each state are made about every 7 to 8 years, which can be a very long time in areas of rapid land-use change. This has led the SOFIA into a dynamic search for intermediate low-cost updating techniques. Recently, the attention has been on the use of Advanced Very High Resolution Radiometer (AVHRR) data for detection of area change. These efforts by SOFIA are described in Zhu (1992) and Zhu and Teuber (1991). In particular, Zhu and Teuber (1991) found that the large pixel size of AVHRR data (1.1 km square) contributed to bias in the percent forest cover estimates. This is a common problem when the scale of measurement differs substantially from the scale of interest or definition. The coarseness of the data contributes not only to excessively smoothed edges but also to both missed non-forest enclaves in forest land and missed forest exclaves in non-forest land. Calibration by smaller scale measurements should greatly improve the accuracy of forest cover estimates, especially in highly fragmented land-use areas.

In order to provide updated estimates of percent forest cover, a method of adjusting the coarse-grained estimates from AVHRR data with estimates from a small finer-grained sample, such as from Landsat Thematic Mapper (TM) data (30 m by 30 m) or field plots, is necessary. The finding by Zhu (1992) that the AVHRR based estimates differed from the accepted Forest Inventory and Analysis (FIA) estimates inspired this work, in which we calibrate the AVHRR estimates with FIA estimates from a small sample of counties using ratio and regression estimators. The procedure assumes that the FIA estimate is correct or at least the one to predict. We adjust the AVHRR estimates with a few FIA estimates in an attempt to get closer to the answer we would have gotten had we done a complete FIA survey. We also investigate the inclusion of population density as a surrogate for land-use fragmentation, which Zhu and Teuber (1991) found to contribute to differences in the two estimates. Both the ratio and regression estimators will be shown to give favorable results with some small gain possible when population density is considered.

Data
The data for this study were provided by SOFIA and consisted of two estimates of percent forest area in 1982 by county for the 67 counties in the state of Alabama. One estimate was obtained from standard FIA plots and one estimate from interpretation of AVHRR data. The processing and supervised maximum-likelihood classification of the AVHRR data is described in Zhu (1992) and Zhu and Teuber (1991). In addition, population by county was estimated by the U.S. Census (Anonymous, 1992).

Estimators

Ratio Estimators
The ratio estimators in this section utilize the AVHRR estimates from all of the counties in the state and paired AVHRR and FIA estimates from a small sample of counties. These estimators assume that the estimate from a complete FIA inven-

F.A. Roesch, Jr., and P.C. Van Deusen are with the Institute for Quantitative Studies, USDA Forest Service, Southern Forest Experiment Station, 701 Loyola Avenue, New Orleans, LA 70113.

Z. Zhu is with the Southern Forest Inventory and Analysis Unit, Forestry Sciences Laboratory, USDA Forest Service, Southern Forest Experiment Station, Starkville, MS 39759.

Photogrammetric Engineering & Remote Sensing, Vol. 61, No. 3, March 1995, pp. 307–311.

0099-1112/95/6103-307$3.00/0

tory could be modeled as some constant proportion of the AVHRR estimate. We consider two ratio estimators resulting from different assumptions on the variance of the FIA estimate over the range of the AVHRR estimates.

Let x_i equal the AVHRR estimate of the percent forest cover in county i, y_i the FIA estimate of the percent forest cover in county i, and a_i the total area of county i, in acres. After plotting y_i against x_i in Figure 1 for all of the counties, we see that the variance of y_i is either close to constant over the range of x_i or it is proportional to $1/x_i$. Each variance assumption leads to a different ratio to be used in the estimator. Of course, y_i would not usually be available for the entire population, in which case the assessment of the variance structure must be based on the sample data alone. Following Cochran (1977, p. 160) for the case of constant variance of y_i for each x_i, we get the ratio

$$\hat{R}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{1}$$

where n is the number of counties for which an FIA estimate is obtained.

If the variance of y_i were instead assumed to be proportional to $1/x_i$, the optimal ratio to use in the ratio estimator would be

$$\hat{R}_2 = \frac{\sum_{i=1}^{n} x_i^2 y_i}{\sum_{i=1}^{n} x_i^3} \tag{2}$$

The two ratio estimators of percent forest land would then be formed: i.e.,

$$Y_{R_j} = \mathbf{W}'\mathbf{X}\hat{R}_j \; ; \; j=1,2 \tag{3}$$

where $\mathbf{W}$ is the $N\times 1$ vector of the relative county areas (w_i); i.e., the county area (a_i) divided by the area of the state A, and $\mathbf{X}$ is the $N\times 1$ vector of x_i values for the entire state.

To obtain a sample estimate for the variance of Y_{R_j}, the residual vectors $\mathbf{r}_j$ can be formed by calculating each of the n elements as

$$r_{ji} = (y_i - \hat{R}_j x_i)w_i; \; j=1,2; \; i=1,\ldots,n. \tag{4}$$

The sample estimate of the variance of each Y_{R_j} is then

$$v(Y_{Rj}) = \frac{N^2\,(1-f)}{n(n-1)}\,\mathbf{r}_j'\mathbf{r}_j; \quad j=1,2 \tag{5}$$

where f is the sampling fraction equal to n/N.

Regression Estimators

The simple linear regression estimator calibrating the AVHRR estimate from the entire state with the small sample of paired AVHRR and FIA estimates is found by first estimating the 2×1 coefficient vector

$$\hat{\mathbf{b}}_1 = (\mathbf{X}_n' \, \mathbf{X}_n)^{-1} \, \mathbf{X}_n' \, \mathbf{y} \tag{6}$$

where

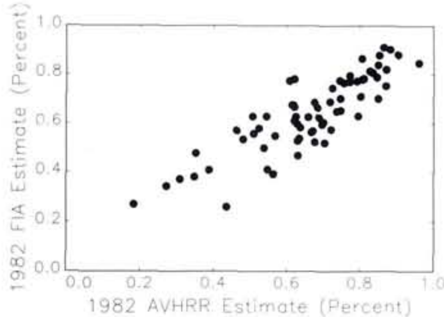

Figure 1. The estimates of percent forest cover for each county of Alabama from a standard FIA inventory verses the same estimates from AVHRR data.

$$\mathbf{X}_n = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \tag{7}$$

The simple linear regression estimator of percent forest land is

$$Y_{LR} = \mathbf{W}'\mathbf{X}_N\hat{\mathbf{b}}_1 \tag{8}$$

where

$$\mathbf{X}_N = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_N \end{bmatrix} \tag{9}$$

Again, as we did for the ratio estimators, after individually calculating the elements of the residual vector ($\tilde{\mathbf{r}}$),

$$\tilde{r}_i = (y_i - (\mathbf{X}_n\hat{\mathbf{b}}_1))w_i, \tag{10}$$

we estimate the variance of Y_{LR} by

$$v(Y_{LR}) = \frac{N^2(1-f)}{n(n-1)}\,\tilde{\mathbf{r}}'\tilde{\mathbf{r}}. \tag{11}$$

We can incorporate our knowledge of the population density into the regression estimator in a number of ways, depending on how we think population density is related to land-use fragmentation and then how fragmentation explains the difference between the two estimates. We chose to use population density as an additional variable in a multiple regression. After a graphical analysis of the data, the natural log (ln) transformation of population density was chosen to be used in the multiple regression, because this transformation appeared the most linear. To facilitate the multiple regression method we could form the matrices for the sample of counties; i.e.,

$$\mathbf{X}_{nn}=\begin{bmatrix}1 & x_1 & p_1\\ 1 & x_2 & p_2\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ 1 & x_n & p_n\end{bmatrix} \tag{12}$$

where p_i is natural logarithm of population density of county i, and for all of the counties,

$$\mathbf{X}_{NN}=\begin{bmatrix}1 & x_1 & p_1\\ 1 & x_2 & p_2\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot\\ 1 & x_N & p_N\end{bmatrix} \tag{13}$$

The estimate of the coefficient vector is found by replacing $\mathbf{X}_n$ with $\mathbf{X}_{nn}$ in Equation 6 to obtain, say, $\hat{\mathbf{b}}_2$. Likewise, replacing $\mathbf{X}_N$ with $\mathbf{X}_{NN}$ and $\hat{\mathbf{b}}_1$ with $\hat{\mathbf{b}}_2$ in Equations 8 and 10 yields the multiple regression estimator of percent forest cover, Y_{MR}, and the elements of its residual vector. This resulting residual vector is then used as it was in Equation 11 to obtain the sample estimate of the variance, $v(Y_{MR})$. Although we will not do so in this paper, additional X variables, such as the corresponding estimates from Landsat TM data, could be added to the regression estimator by appropriately redefining the $\mathbf{X}$ matrices.

Methods

Simulation Description

Each run of the simulation consisted of drawing simple random samples without replacement of the counties in Alabama using sample sizes of 8, 10, 12, 14, 16, 18, and 20. When a county was chosen for the sample the FIA estimate of percent forest cover and the corresponding AVHRR estimate for that county were used. These sample estimates were combined with the statewide estimate from the AVHRR data according to the rules of the two ratio estimators and the two regression estimators to estimate percent forest cover for the state. After 10,000 runs, the mean-squared error (MSE) and bias were estimated for each estimator and each sample size using the FIA estimate from all 67 counties as the "true" percent forest cover for the state. The variance of the 10,000 estimates of each type was also calculated.

The same statistics were also calculated for the variance estimators. The approximate true variances, as given in Cochran (1977, p. 153 and 194), were treated as the "true" variances and used for the MSE and bias calculations of the simulation.

Results

Figure 2 gives the results for the estimators of percent forest cover after 10,000 simulations of the even sample sizes 8 through 20. The figure shows the graphs of MSE, variance, and bias, relative to the statistics for Y_{MR}, as defined above. The lower x-axis graphs the sample size and the upper x-axis gives the value of the statistic for Y_{MR}.

We see that, in general, none of the estimators show any unreasonable level of bias, although the regression estimators have less bias than the ratio estimators. Y_{R_2} displays the most bias, always more than twice that of Y_{MR}. Except at a sample size of 8, Y_{LR} shows the least bias.

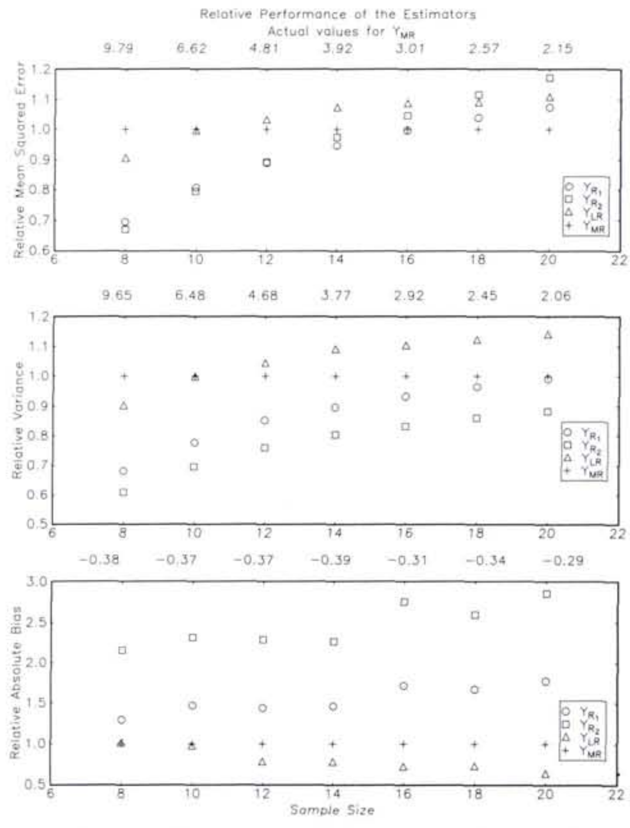

Figure 2. The relative mean-squared error, relative variance, and relative absolute bias of the estimators of percent forest cover verses sample size after 10,000 runs of the simulation. These values are relative to the respective statistics for Y_{MR} at each sample size, which are explicitly given above the upper x axis. The percent forest cover from the FIA data, the "true" value, was 65.16.

The ratio estimators are clearly superior in terms of variance at all sample sizes tested, with Y_{R_2} having the lowest variance throughout, supporting the assumption that the variance of y_i over the range of x_i is approximately proportional to $1/x_i$. The regression estimator utilizing the population density information (Y_{MR}) dominates the simple linear regression estimator (Y_{LR}) at all sample sizes greater than 10.

The ratio estimators are superior in terms of mean-squared error at the smaller sample sizes, but the influence of bias becomes more important for the ratio estimators as sample size increases and variance decreases. Y_{MR} is the lowest in MSE at the two largest sample sizes, with Y_{R_1} holding second place.

Figure 3 gives the results for the variance estimators after the 10,000 simulations. The figure graphs the relative MSE, relative variance, relative absolute bias, and mean of the variance estimators. As in Figure 2, the lower x-axis gives the sample size and the upper x-axis of the upper three graphs gives the value of the statistic for $v(Y_{MR})$. This figure shows that both the variance and MSE of $v(Y_{MR})$ are noticeably smaller than those of the other variance estimators at all

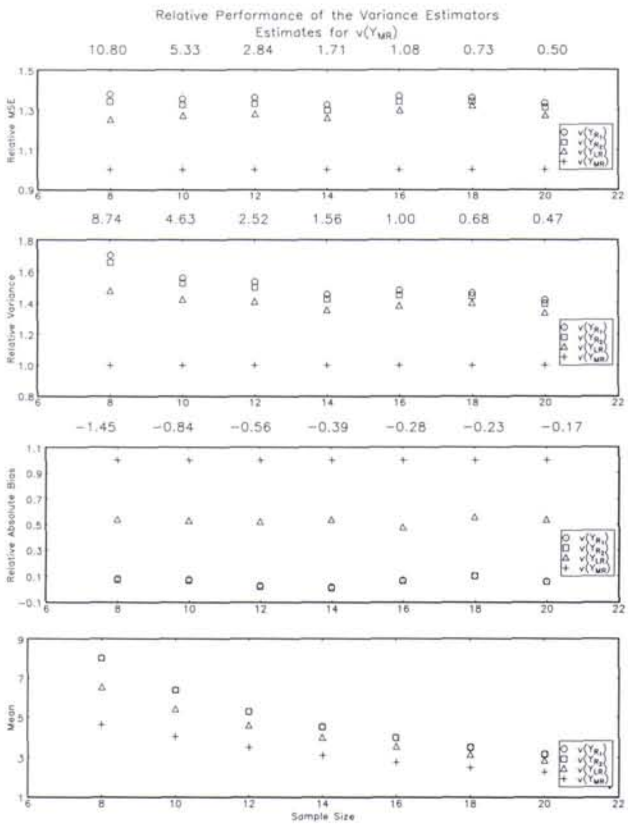

Figure 3. The relative mean-squared error, relative variance, relative absolute bias, and mean of the variance estimators verses sample size after 10,000 runs of the simulation. The top three values are relative to the respective statistics for $v(Y_{MR})$, which are explicitly given above the upper x axis.

sample sizes. Second and third places for these statistics at all sample sizes are held by $v(Y_{LR})$ and $v(Y_{R_2})$, respectively, although the differences are much smaller. The bias of $v(Y_{MR})$ is an order of magnitude greater than the biases of $v(Y_{R_1})$ and $v(Y_{R_2})$. The bias throughout, however, isn't large enough to effect a change in ranking of the estimators between the variance and mean squared error plots.

Conclusions

Zhu (1992) had shown that the use of AVHRR data can result in a valuable updating tool. We have shown that a reasonable improvement in AVHRR updates of percent forest area and change in forest area could be obtained by incorporating a small number of county level FIA estimates through both the ratio and regression estimators.

Very often, sample size is determined by logistical and financial constraints rather than by statistical constraints. If one were interested in using the smallest sample of counties possible in order to minimize cost or field time, the ratio estimators would be hard to beat. These results correspond almost exactly to those of Rao (1969), as quoted in Cochran (1977), in which eight natural populations were included in a simulation study. That study demonstrated empirically that the dominance of the linear regression estimator over the ratio estimator in terms of MSE is only applicable to large samples. Rao (1969) found that the average of the ratios of MSE for the linear regression estimator over the ratio estimator was 1.36 for a sample size of 8 and 1.15 for a sample size of 12.

Besides comparing the estimators with each other, we must also consider the gain achieved by using any of the estimators. Over the entire state, the squared difference between the FIA estimate and the AVHRR estimate is 5.37. Therefore, if an estimator in our simulation did not produce an MSE of less than 5.37, the resulting estimates would have, on the average, been farther from the "truth," by the squared-error criterion, than the AVHRR estimate. In this study, larger MSEs than this were obtained with sample sizes less than 10, while a sample size of 12 was required for the regression estimators to fall below this threshold. FIA estimates on at least 12 counties were necessary in this instance to achieve a notable reduction in MSE.

It is possible that some other easily obtained but more direct measure of land-use heterogeneity would further improve estimates of percent forest land and possibly allow the use of fewer surveyed counties. County level estimates from higher resolution LANDSAT data could be used in place of the FIA estimates in the estimators above for a possible further reduction in cost of calibrating the AVHRR data.

References

Anonymous, 1992. *Rand McNally Road Atlas, 68th Edition,* Rand McNally, Chicago, Illinois, 136 p.

Cochran, W.G., 1977. *Sampling Techniques, 3rd Edition,* John Wiley & Sons, New York, 428 p.

Rao, J.N.K., 1969. Ratio and regression estimators, *New Developments in Survey Sampling* (N.L. Johnson and H. Smith Jr., editors), John Wiley & Sons, New York, pp. 213-234.

Zhu, Z., 1992. *Advanced Very High Resolution Radiometer Data to Update Forest Area Change for Midsouth States,* Res. Pap. SO-270, USDA, Forest Service, Southern Forest Experiment Station, New Orleans, Louisiana, 11 p.

Zhu, Z., and K.B. Teuber, 1991. AVHRR data for forest area change in Alabama, *Proceedings of ACSM-ASPRS Annual Convention,* 25-29 March, Baltimore, Maryland, 3:516-522.

(Received 10 August 1992; revised and accepted 23 June 1993)

Francis A. Roesch, Jr.

Francis A. Roesch, Jr. received the B.S. degree in Natural Resource Management in 1984, the M.S. degree in Statistics in 1988, and the Ph.D. degree in Forest Biometrics in 1988, all from Rutgers University. Since then he has been at the Institute for Quantitative Studies in the USDA Forest Service's Southern Forest Experiment Station. A unifying theme to his work concerns the development of sampling tools for the timely description of forests and change in forest ecosystems.

Paul C. Van Deusen

Paul C. Van Deusen received the B.S. degree in Forest Management in 1975 from the University of Massachusetts, the M.S. degree in Forest Biometrics in 1979 from Mississippi State University, and the Ph.D. degree in Forest Biometrics in 1984 from the University of California. He is currently the Project Leader at the Institute for Quantitative Studies in the USDA Forest Service's Southern Forest Experiment Station.

Zhiliang Zhu

Zhiliang Zhu is a Research Forester with the Southern Forest Experiment Station's Forest Inventory and Analysis Unit, USDA Forest Service, in Starkville, Mississippi. He received both the M.S. and Ph.D. degrees in Natural Resources from the University of Michigan. His current research involves large area forest mapping (most recently in boreal and Central American regions), and developments of remote sensing procedures in support of forest inventories in the mid-south United States.