Analytical Alternatives for an Annual Inventory System

Methods for analyzing data from the Southern Annual Forest Inventory System (SAFIS) are discussed. Differences between the annual inventory approach and the more traditional periodic approach require that we revisit the previous assumption that there are no important spatial and temporal trends in the data. Over the next few years, the USDA Forest Service Southern Research Station will be evaluating models of varying complexity to determine the most efficient estimation approach for each variable, at all spatiotemporal scales of interest.

By Francis A. Roesch and Gregory A. Reams

The USDA Forest Service Southern Research Station (SRS) Forest Inventory and Analysis Unit (FIA) has initiated an annualized forest inventory sampling design, the Southern Annual Forest Inventory System (SAFIS). SAFIS was introduced to improve estimation of both the current resource inventory and changes in the resource. Under the previous periodic inventory system, individual states were inventoried over a two- to three-year period, about every 10 years. Many factors, including rapid land use changes and the intense forest dynamics in the southern United States, contributed to diminished confidence in inventory estimates that were more than a few years old. It was decided that an annualized inventory system, in which data is collected statewide every year, would provide more timely and useful estimates. We will discuss some of the analytical proposals for data from this system.

Before the SAFIS effort, the North Central Research Station (NCRS) had been conducting an annualized inventory in Minnesota in cooperation with the Minnesota Department of Natural Resources. More recently, the Agricultural Research, Extension, and Education Reform Act of 1998 (PL 105-185) directed the entire Forest Service to move toward an annualized inventory. Although this article addresses SAFIS directly, recent developments have led to SRS and NCRS scientists joining forces to investigate the challenges and opportunities arising from this transition to annual inventories.

The plot arrangement for the SAFIS sample design resulted from an intensification of the National Forest Health Monitoring (FHM) grid, which has been described as a component of a global environmental monitoring sample design (Overton et al. 1990; White et al. 1992). The sample plots are located in a systematic triangular grid with five interpenetrating panels. One panel per year is measured for five consecutive years. Every five years the panel measurement sequence reinitiates. If panel 1 was measured in 1998, it will also be measured in 2003, 2008, and so on. Panel 2 would then be measured in 1999, 2004, 2009, and so on. The panels will be as well dispersed as possible if we apply them according to the pattern in *figure 1*. Note that in a triangular grid the cells are hexagonal in shape. The result of this pattern is that each element has no immediate neighbors from the same panel.

Implementation of SAFIS requires a transition from one of two variations of a periodic system to the rotating panel design described above. The first of the two variants of the periodic sample design, that found in the western states within the SRS area of responsibility, consists of a collection of three-square-mile grids placed randomly within each survey unit. The survey units are of such a size that there are typically

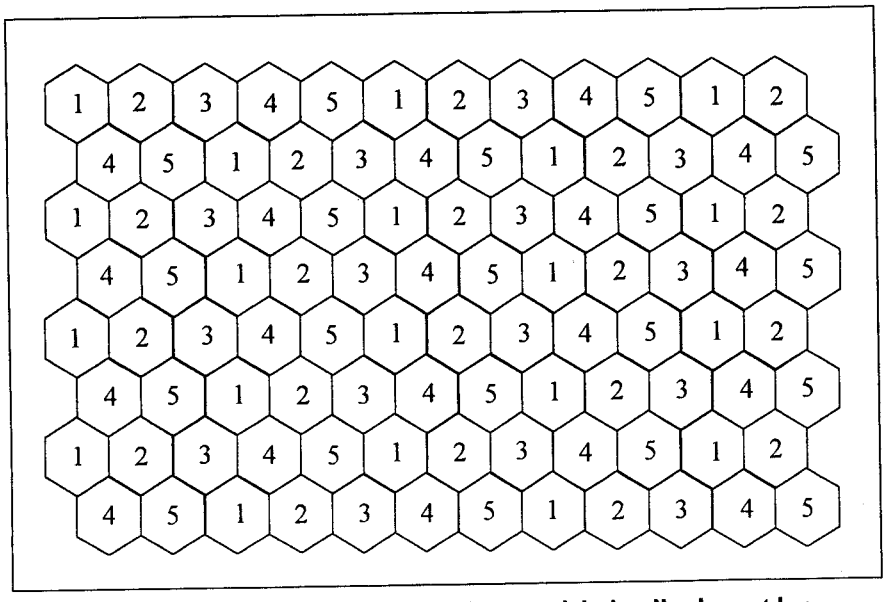

Figure 1. **An interpenetrating pattern for a five-panel design. No element has another member from the same panel as an immediate neighbor.**

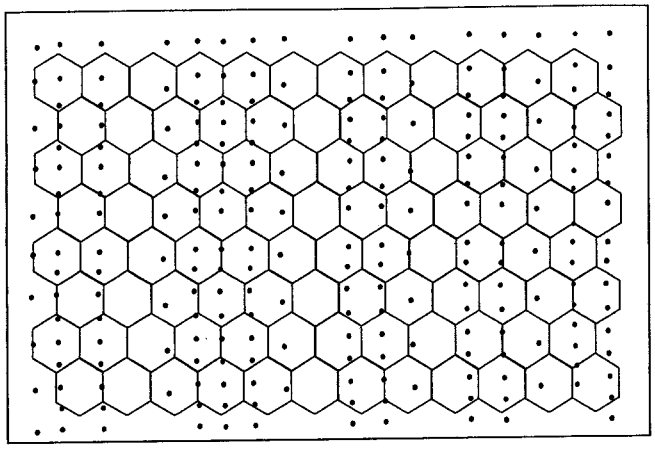

Figure 2. **Example of a coarse mapping of existing plot locations.**

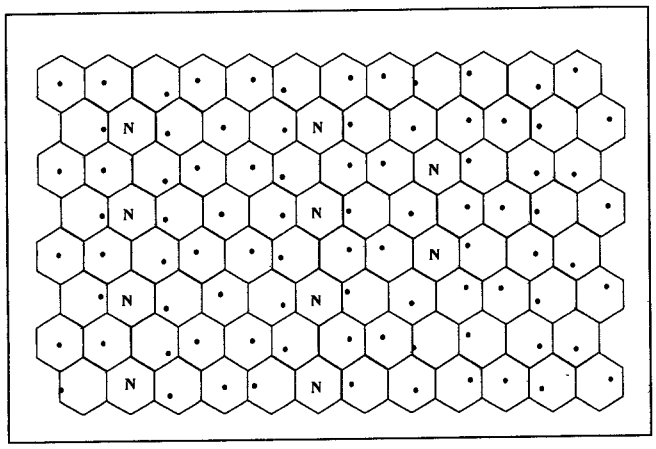

Figure 3. **Example of a coarse mapping of existing plot locations after deletions of unneeded plots. Cells containing an "N" would require new plots under option (3a).**

several within a state. The second variant of the periodic sample design, occurring in the eastern SRS states, underwent a number of changes over past decades. Unfortunately, not all of these changes have been well documented. The resulting pattern of plot locations on the landscape is a somewhat irregular grid, of a higher spatial density than desired.

SAFIS has presented a few challenges with respect to inventory goals that can sometimes be at odds. For instance, an obvious goal would be to ensure that the transition is as smooth as possible, and another goal would be to implement the new design as quickly as possible while minimizing cost. The goal of a smooth transition can conflict with the goal of quickly implementing the new design. To ensure a smooth transition, we must maintain temporal consistency and continuity for trend estimation. We could argue that this would be easier to accomplish if we retained as many of the old plot locations as possible. Although it is true that there is a cost associated with establishing new field plot locations, the quickest and easiest implementation of the new design would occur if an entirely new grid of sample points is established across the SRS area of responsibility.

There are numerous methods we might use to choose existing sample point locations for retention in the new design. The different methods involve varying degrees of compromise between simplicity and the desire to sacrifice as little of the historical trend information as possible. The options we consider here are to:

1. eliminate all the old plot locations and start over with a triangular grid
2. delete plot locations until a roughly regular grid results, at the same intensity as the desired grid
3. use a coarse mapping to assign existing plots to the nearest grid point *(fig. 2)*. Subsequent to the coarse mapping, we could:
   a. delete any extra plots in each grid cell and establish new plots at the center of every empty cell *(fig. 3)*, or
   b. assign residual plots within one grid cell of an empty cell to the empty cell and establish new plots at the center of any still-empty cell *(fig. 4)*.

Option 1 is the most expensive and option 2 is the least expensive. The variations of option 3 are considered a compromise because they do not result in a regular grid of points at a fine scale, but they do at a coarse scale. Also, they provide a formal mechanism for assigning the existing locations to a regular grid of cells, and could be analyzed, with caveats, as though the sample consisted of a regular grid.

The advantage of option 1 is that we would be starting fresh with nothing messy or complex to compensate for in the future. Also, the entire SRS would operate under a single sample design with no conflicts in concept or analytical procedure. The disadvantages are that it would result in a gap in observations of trend and would necessitate replication of all pre-field work with respect to point location identification and classification.

Option 2 has the advantage of being quick and easy to implement. In addition, trend information will benefit from the continuity of plot locations. The drawbacks include the fact that some clumping of plot locations will occur and, if spatial relationships are modeled, there could be potentially large differences in the analytical procedures between eastern and western states for some variables.

Option 3a also has the advantages that trend information will benefit from the continuity of sample locations and some of the work that goes into point location identification and classification will be reusable. Less clumping of plot locations will occur than with option 2 and there will be at most small differences in analytical procedures within the station. As with option 2, the actual location of plots would not be regularly spaced at the finest scales of measurement. We note that option 3b retains more of the original plot locations than option 3a. Therefore, trend information will benefit to a greater extent under option 3b. Most of the work that has gone into point location identification and classification will be reused. Again, as with option 3a, small differences of analytical procedures may be necessary.

Analysis

As we discuss the different analytical approaches for the SAFIS design, we assume that option 3a above will be used to assign existing plots to their enclosing cells and that new plots will be established at the centers of all empty cells. We

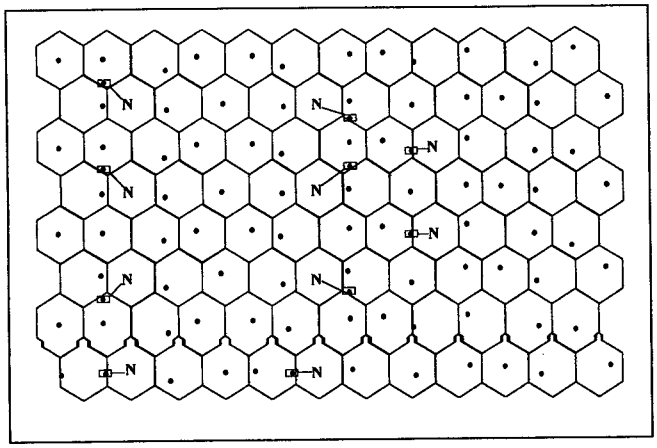

Figure 4. **Option (3b) would utilize extra plots in adjacent cells for assignment to empty cells.**

recognize that there will be some demand for analyses during years 1 through 4 of the annual design in a given area. For these transition years, FIA will provide composite estimates formed by the appropriate weighting of estimates from the complete periodic inventory combined with estimates from the incomplete annual inventory. In this article, we examine procedures to be used once a full series of observations is available; that is, when all five panels have been measured at least once.

The major difference among the analytical procedures being proposed is in the extent to which spatial patterns and time trends are ignored. That is, in extensive inventories, such as SAFIS, one may or may not wish to make the usual assumption that spatial location or actual year of measurement within a single panel series is unimportant in the analysis. We will briefly discuss some of the analytical procedures being proposed while keeping this perspective at the forefront. To do this we must provide notation that will allow the use of a full spatial-temporal model; however, we will occasionally be able to collapse the model along one or more dimensions.

One of the measured variables for each plot will be the proportion of plot area in each condition class. A condition class is defined as the combination of variables that identify different strata. Forest condition classes are at least an acre in size and identified by land use, forest type, stand origin, stand size, stand density, and ownership class (Anonymous 1998). Assume that we seek estimates for each condition class observed in a survey unit. Also assume that the horizontal and vertical positions of cell centers are numbered from west to east and south to north, respectively. Let:

h_i = horizontal position i ($i = 1,\ldots,I$)
v_j = horizontal position j ($j = 1,\ldots,J$)
t_t = time t ($t = 1,\ldots,5$)
c_k = condition class k ($k = 1,\ldots,K$)
X_{ijtk} = the per-acre value observed at h_i, v_j, and t_t, for c_k
A_{ijtk} = the area in acres sampled in c_k at h_i, v_j, and t_t

$$C_{ijtk} = \begin{cases} 1 & \text{If } c_k \text{ occurs at } h_i, v_j, \text{ and } t_t, \\ 0 & \text{Otherwise} \end{cases}$$

A_P = plot area

The focus here is on estimation of a per acre value (V) for condition class k under different assumptions of spatial and temporal trend.

If we assume that there is no time or spatial trend at the observed scales, then our data model would have the simplest form possible, and the overall mean for the five-panel series would provide the best estimator of a per-acre value (V) for condition class k:

$$V_k = \frac{1}{A_{S_k}} \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{A_{ijtk}}{A_P} X_{ijtk}$$

Otherwise, if we were willing to ignore any spatial trend, we could calculate the mean within each panel for an estimate each year:

$$V_{tk} = \frac{1}{A_{S_{tk}}} \sum_{i=1}^{H} \sum_{j=1}^{J} \frac{A_{ijtk}}{A_P} X_{ijtk}$$

where $A_{S_{tk}}$ = sum of the plot areas sampled in condition class k at time t.

This approach would, however, provide an inadequate sample for many variables. Rather, we should explore different models for the time trend to efficiently use the entire five-panel sample. The simplest model, that of no time trend, would weight the panels equally:

$$V_k = \sum_{t=1}^{5} .2 V_{tk} \qquad (1)$$

This is the method used by FIA for periodic inventories in states that required more than one year to inventory. An advantage to using this approach initially is that the current software used by FIA would be applicable.

Because the time duration of measuring all five panels is somewhat longer than the duration of one to three years per state that it took for the periodic inventories, equal weighting of plots across panels may have the tendency to mask temporal trends. One suggested solution for this problem has been to form an estimator in which panels that were-measured more recently are weighted more heavily than those measured earlier. This estimator would take the following form:

$$V_k = \sum_{t=1}^{5} w_t V_{tk} \qquad (2)$$

And w_1 through w_5 would be weighted so that the sum is equal to 1, such as .1, .1, .2, .3, and .3, respectively. If one used the preceding weighting scheme, it would be analogous to stating that one has three times as much confidence in panels 4 and 5 being fair representations of today's condition as panels 1 and 2.

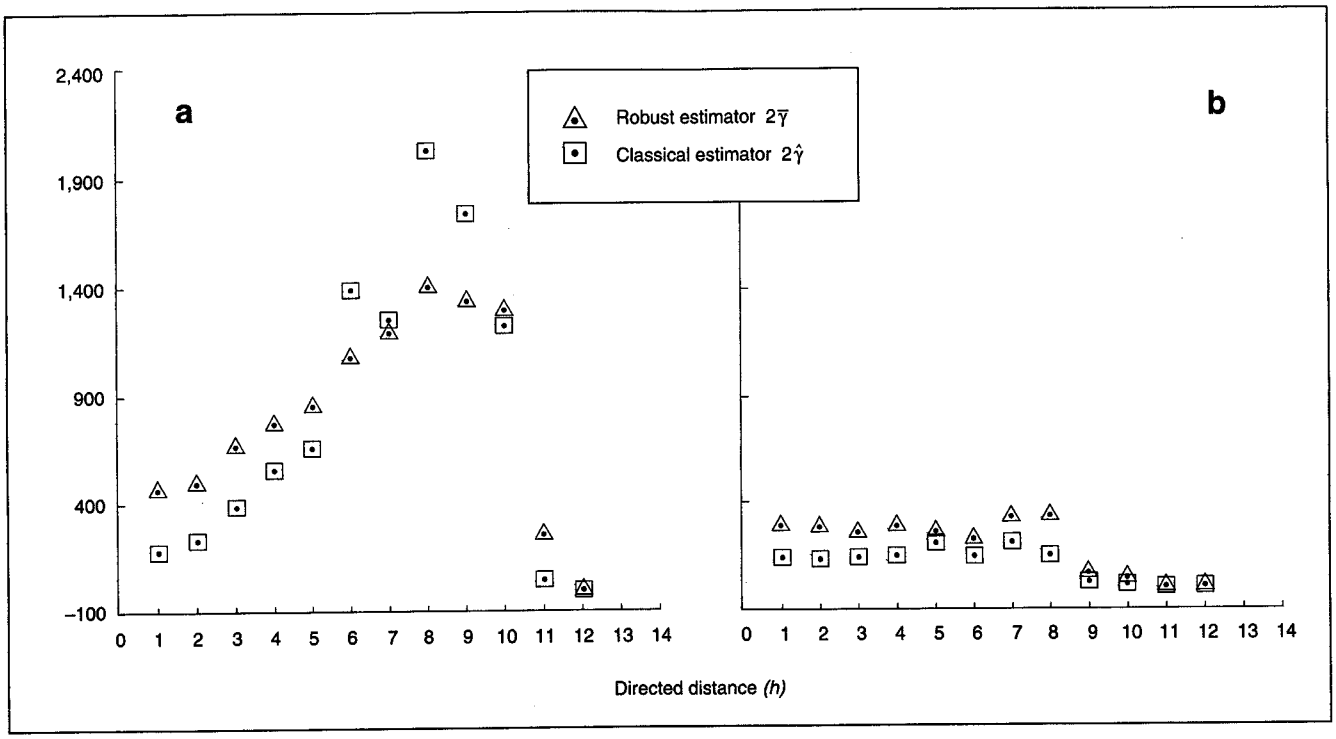

Figure 5. **(a) Estimated NE/SW variogram for percent basal area spruce/fir. (b) Estimated NE/SW variogram of residuals from the median polish.**

A less arbitrary approach would be to attempt to model the time trend within a panel series. Van Deusen (in review) presents a mixed estimator that can incorporate increasing levels of constraints on the derivatives of the time trend, allowing one to model various levels of complexity in the time trend. The mixed estimator literally mixes two models: the first describes the relationship of observations within each panel (or time period) and the second describes the time trend. The mixed estimation approach is both powerful and practical for most variables of interest to FIA. A slightly more complex formulation than that given by Van Deusen would be appropriate to satisfy FIA's charge to recognize changes in condition class within field plots. This might be considered necessary because, although each plot samples the same amount of surface area, the area of a condition class sampled by a plot can vary from zero to the size of a plot. Therefore, individual plot averages for a particular condition class have different bases of support and should probably be weighted accordingly.

Finally, as pointed out in Roesch (1994) for the case of forest health monitoring, some variables will display spatial trends within condition classes in extensive inventories. Spatial analyses are of interest any time the measurement of a variable is likely to be different solely because of the spatial location of the observation. In these cases, a larger class of models, which include spatial correlation, should be used. Along these lines we could fully analyze the effect of all of the spatial dimensions or we could implicitly undermine the importance of one (or more) of the dimensions by collapsing it down into the remaining dimensions (as is usually done for elevation). Having the ability to discover and remove spatial correlation makes it easier to investigate other potentially important relationships in the data, and will at times provide a simple explanation for high variability in a measurement of interest.

To perform a spatial analysis at a particular scale, the first step is usually a coarse mapping, which is accomplished by segmenting an area with a specific size grid and pooling the plots within each segment (see Cressie 1991). Next, the median polish technique is often used to decompose the value in each cell at each time (X_{ijtk}) into its assumed components of an effect common to all cells, spatial effects in two directions, and a residual:

$$X_{ijtk} = C_{tk} + H_{itk} + V_{jtk} + R_{ijtk}$$

where:

C_{tk} = the "common" effect at time t ($t=1,\ldots,5$)
H_{itk} = the ith horizontal effect ($i=1,\ldots,I$)
V_{jtk} = the jth vertical effect ($j=1,\ldots,K$)
R_{ijtk} = the residual in cell i,j at time t

For T time periods, the result is a $1 \times T$ vector C of common effects, two matrices (H and V) of directional effects, and a matrix R of residuals.

Although there are ways other than the median polish to accomplish this decomposition, we do not want the effects to be overly influenced by any outliers present. See Cressie (1991) for a defense of the two-way median polish when outliers are a potential concern. The matrix R can be evaluated for special cases in the same manner as the more familiar residual analysis for regression. Subsequent to the median polish we can obtain residuals that are not time-detrended by adding the "common" effect for each time period back

into the residuals:

$$W_{ijtk} = R_{ijtk} + C_{tk}$$

We could then treat the matrix W as an independent set of time-series observations, and analyze the time trend.

One way to ensure that suspected spatial trends have been removed is to estimate the variogram at a series of directed distances. The variogram is the variance of the difference in values, separated by a specific distance and direction, observed at defined points in space. If s represents an observation point, h represents a directed distance, and $X(s)$ represents the value of the variable at point s, then the variogram is defined as $2\gamma(h) = \mathrm{var}(X(s+h) - X(s))$. By plotting estimates of the variogram for different values of h, we can determine the magnitude of spatial correlation for a variable at different scales.

The classical estimator of the variogram is:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (x_i - x_{(i+h)})^2$$

where:

$N(h)$ = the number of distinct pairs of points separated by directed distance h

$x_{(i+h)}$ = the estimate (or observation) of the variable at the point separated from point i by directed distance h

Any trend in variogram estimates will show the spatial correlation in the variable. For example, *figure 5*, taken from Roesch (1994), shows the variogram plots before and after a median polish was used to remove spatial correlation from the data. The relatively flat variogram of *figure 5(b)* shows that the median polish had effectively removed the spatial correlation from the data represented by the variogram in *figure 5(a)*.

Conclusion

This discussion of the proposed methods for analyzing data from the SAFIS design has shown how these proposals differ mostly in the level of simplification accepted. Traditionally, FIA has given estimates for survey units, which are fairly extensive areas of land within a state. Plots in a survey unit were measured in one or at most two years. Therefore, it was not only reasonable but necessary to ignore time trend in variables during the execution of the survey. In addition, the survey units were thought to be small enough that spatial trend within the unit was not important for the variables of interest. The new sample design has two profound effects: the importance of the survey unit as a logistical tool is eliminated, and the measurement of plots is spread out over five years. These effects require that we revisit the previous assumptions of there not being important spatial or temporal trends within an area for each variable of interest. Probably we will find that these assumptions are often appropriate and, in these cases, the use of the estimator in equation (1) will be valid. Over the next few years, FIA will be evaluating models of varying complexity to determine the most efficient estimation approach for each variable, at each spatial scale of interest.

Literature Cited

ANONYMOUS. 1998. *Field instructions for southern forest inventory, version #2.* Asheville, NC: USDA Forest Service, Southern Research Station.

CRESSIE, N. 1991: *Statistics for spatial data.* New York: John Wiley & Sons.

OVERTON, W.S., D. WHITE, and D.L. STEVENS. 1990. *Design report for EMAP (Environmental Monitoring and Assessment Program).* EPA/600/3-91/053. Washington, DC: US Environmental Protection Agency, Office of Research and Development.

ROESCH, F.A. 1994. Spatial analysis for monitoring forest health. In *Proceedings of the section on statistics and the environment of the American Statistical Association,* 104–9. 1994 Joint Statistical Meetings, August 14–18, 1994, Toronto.

VAN DEUSEN, P.C. In review. Modeling trends with annual survey data.

WHITE, D., J. KIMBERLING, and S.W. OVERTON. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems* 19:5–22.

Francis A. Roesch (e-mail: froesch/srs_fia@fs.fed.us) is mathematical statistician, and Gregory A. Reams is section head, Remote Sensing and Statistical Techniques, USDA Forest Service, Southern Research Station, 160A Zillicoa Street, PO Box 2750, Asheville, NC 28802.