The Effects of Removing Condition Boundaries on FIA Estimates

David Gartner and Gregory Reams[1]

Abstract. —When Forest Inventory and Analysis (FIA) changed to the national standards for the inventory system, plots with multiple condition codes were introduced to the Southern Station's FIA unit. FIA maps up to five different conditions on completely or partially forested 1/24-acre subplots. This change has made producing inventory estimates more complex because the data are analyzed by condition classes (partial plot) rather than on a whole plot. Methods for analyzing by condition classes are less intuitively obvious than methods based on a single condition for an entire plot. We compared the current standard of fully mapped plots to two methods that reduce the number of mapped conditions per plot using the following sets of rules. Rule 1 assigns the predominant condition to the entire plot, including plots that are partially forested. Rule 2 maps a single nonforest condition and a single forest condition, with the single forest condition predominant. The effects of these changes were shown by calculating forest area by forest type group and ownership and volume by species and diameter class. The effect of using just one condition per plot (rule 1) increased the estimated total forested area by 0.4 percent. Using rule 1 decreased the calculated total volume for the state by 3.9 percent. Using only one forested condition per plot (rule 2) did not change the total area estimate or the volume estimates. Both methods decreased the estimated variance for the total volume by 12 to 16 percent. The percent changes in the estimated values were greatest in the least occurring table entries, e.g., the rarest combinations of forest type and ownership group or combinations of species and diameter class.

When Forest Inventory and Analysis (FIA) changed to the national standards for the inventory system, plots with multiple condition codes were introduced to the Southern Station's FIA unit (Reams and Van Deusen 1999). These multiple conditions per plot require determining the boundary between the conditions in the field. In natural stands, forest type boundaries are frequently a continuum, as opposed to abruptly changing from one forest type to another. Therefore, determining the boundary between two forest conditions can be difficult and nonrepeatable. This report addresses two possible changes to the current procedures. One would be to ignore all the boundaries and just use the largest condition on the plot (rule 1). The other possible change would be to ignore boundaries between forested conditions, and use one (the largest) forested condition per plot and one nonforested condition. Our study sought to determine the effects of these possible changes on State estimates of area and volume.

Methods

To simulate the removal of all boundaries, the condition variables, such as forest type and stand age, of the largest condition were assigned to the entire plot. To simulate the possible removal of boundaries between forested conditions, the condition variables of the largest forest were attributed to all the forested area on the plot. For those plots where the two largest forests were the same size, one of the two was chosen at random to be attributed to the larger area, whole plot, or total forested area on the plot. The data came from South Carolina's first panel.

Each plot (sample unit) consists of four 1/24-acre subplots. The condition boundaries actually occur on the subplots. However, there were not sufficient data to analyze the two alternative methods at the subplot level in the database for South Carolina's first panel.

The table values were calculated in the standard manner (FIA Stat Band, in review), except for using double sampling for forested area estimates. First, the phase 1 photo interpretation data were used to estimate the amount of forested land per county. The phase 1 photo interpretation plots are ground truthed on phase 2 plots and a set of ground truth intensification plots. Because condition data are only gathered on the

[1]Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN 37919. Phone: 865–862–2066; e-mail: dgartner@fs.fed.us, and Project Leader, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN 37919.

done

phase 2 plots, the two possible rule changes could not be applied to the photo interpretation intensification plots.

The phase 2 plot data were used to post-stratify the forested area by forest management type. The table values were then calculated using post-stratified sampling estimation techniques (Cochran 1977). To be able to detect any differences between the effect on estimating area and the effect on estimating volume, one area table and one volume table were calculated for each method of handling boundaries. The two tables are 'Table 2—Area of forest land by forest-type group and ownership class' and 'Table 19—Volume of live trees on timberland by species and diameter class' (Thompson 1998).

Specifically, the values for the area table were calculated using equation 1.

$$\hat{Y}_{ot} = A_{total} * \hat{P}'_f * \hat{R}_{ot} \tag{1}$$

where $\hat{Y}_{ot}$ is the estimate of the area for each owner group (o) and forest type (t) combination for a given county, A_{total} is the total land area in the county obtained from the U.S. Census Bureau, $\hat{P}'_f$ is the ratio of forested land to total land area in the county as estimated from the phase 1 data, and $\hat{R}_{ot}$ is the estimated proportion of forest land by ownership group and forest type in the county.

$\hat{P}'_f$ is estimated using double sampling techniques using equation 2 (Reams 2000).

$$\hat{P}'_f = (\hat{P}_f x C_f) + (\hat{P}_n x C_n), \tag{2}$$

where $\hat{P}'_f$ is the proportion of phase 1 plots photo interpreted to be forested, C_f is the proportion of ground truthed plots whose initial photo interpretation calls were forested and were ground truthed as forested, $\hat{P}_n$ is the proportion of phase 1 plots photo interpreted to be nonforested, and C_n is the proportion of ground truthed plots whose initial photo interpretation calls were nonforested that were ground truthed as forested.

$\hat{R}_{ot}$ is the proportion of forested area in a given ownership group (o) and forest type (t) combination as estimated by a ratio of means estimator using equation 3 (Cochran 1977, Zarnoch and Bechtold 2000).

$$\hat{R}_{ot} = \frac{\sum_{i=1}^{n_f} A_{ot_i}}{\sum_{i=1}^{n_f} A_i}, \tag{3}$$

where A_{ot_i} is the area in ownership group by forest type combination on phase 2 plot i, A_i is the amount of forested area on plot i, and n_f is the number of phase 2 plots with at least one forested condition.

Because $\hat{Y}_{ot}$ is a product of two random variables, the variance of $\hat{Y}_{ot}$ (Var($\hat{Y}_{ot}$)) contains the product of the variances of the two random variables and the cross products of those variances and the squared means of the random variables using equation 4 (Goodman 1960, McCollum 2002).

$$Var(\hat{Y}_{ot}) = A_{total}^2 * \{[\hat{P}'_f{}^2 * Var(\hat{R}_{ot})] + [Var(\hat{P}'_f) * \hat{R}_{ot}{}^2] - [Var(\hat{P}'_f) * Var(\hat{R}_{ot})]\} \tag{4}$$

where $\hat{Y}_{ot}$, A_{total}, $\hat{P}'_f$, and $\hat{R}_{ot}$ have the same definitions as equation 1. The variances of $\hat{P}'_f$ and $\hat{R}_{ot}$ are shown in equations 5 and 6.

$$Var(\hat{P}'_f) = \frac{\hat{P}_f * \hat{P}_n}{n} * (C_f - C_n)^2 + [\frac{(\hat{P}_f)^2 * C_f * (1 - C_f)}{m_1}] + [\frac{(\hat{P}_n)^2 * C_n * (1 - C_n)}{m_2}] \tag{5}$$

where $\hat{P}'_f$, $\hat{P}_f$, $\hat{P}_n$, C_f, and C_n have the same definitions as in equation 2, n is the total number of photo interpretation points, m_1 is the number of photo interpretation plots that were ground truthed as forested, and m_2 is the number of photo interpretation plots that were ground truthed as nonforested.

$$Var(\hat{R}_{ot}) = \frac{1}{n_f * (n_f - 1) * [\sum_{i=1}^{n_f} A_i / n_f]^2} * [\sum_{i=1}^{n_f} A_{ot_i}{}^2 + \hat{R}_{ot}{}^2 \sum_{i=1}^{n_f} A_i{}^2 - 2 * \hat{R}_{ot} \sum_{i=1}^{n_f} (A_i * A_{ot_i})] \tag{6}$$

where $\hat{R}_{ot}$, A_{ot_i}, A_i, and n_f have the same definitions as in equation 3.

The values for the volume table for combinations of species group (s) by diameter class (d) are calculated using equation 7.

$$\hat{V}_{sd} = A_{total} * \hat{P}'_f * \sum_{t=1}^{T} \hat{R}_t * \hat{v}_{tsd}, \tag{7}$$

where A_{total} and $\hat{P}'_f$ are as defined in equation 1, $\hat{R}_t$ is the proportion of forested area in a particular forest type, and $\hat{v}_{tsd}$ is the volume per acre for combinations of species groups and diameter classes. Both $\hat{R}_t$ and $\hat{v}_{tsd}$ are ratios of means estimators of the general form found in equation 3 and variances of the general form found in equation 6. $\hat{R}_t$ is the ratio of the total area in forest type t divided by the total forested area summed over plots that contain at least one forested condition. $\hat{v}_{tsd}$ is the ratio of the total volume in the species group by diameter class for forest type t divided by the total area in forest type t summed over plots that contain at least one condition with that forest type.

The variance of $\hat{V}_{sd}$ is calculated using Goodman's formula for the variance of the product of random variables, as with the variance of $\hat{Y}_{ot}$. However, instead of being a product of two random variables, $\hat{V}_{sd}$ is the product of three random variables. Applying Goodman's formula in succession gives rise to the formula for the variance of $\hat{V}_{sd}$ (equation 8).

$$
\begin{aligned}
&Var(\hat{V}_{sd}) = \\
&A_{total}^2 * \{\hat{P}'^{\,2}_f * \sum_{t=1}^{T} Var(\hat{R}_t) * \hat{v}_{tsd}^2 + Var(\hat{P}'_f) * \sum_{t=1}^{T} \hat{R}_t^2 * \hat{v}_{tsd}^2 - Var(\hat{P}'_f) * \sum_{t=1}^{T} Var(\hat{R}_t) * \hat{v}_{tsd}^2 - \\
&\hat{P}'^{\,2}_f * \sum_{t=1}^{T} \hat{R}_t^2 * Var(\hat{v}_{tsd}) - \hat{P}'^{\,2}_f * \sum_{t=1}^{T} Var(\hat{R}_t) * Var(\hat{v}_{tsd}) + \\
&Var(\hat{P}'_f) * \sum_{t=1}^{T} \hat{R}_t^2 * Var(\hat{v}_{tsd}) + Var(\hat{P}'_f) * \sum_{t=1}^{T} Var(\hat{R}_t) * Var(\hat{v}_{tsd})\}
\end{aligned}
\tag{8}
$$

where $\hat{V}_{sd}$ is defined in equation 7, A_{total} and $\hat{P}'_f$ are as defined in equation 1, $\hat{R}_t$ is the proportion of forested area in a particular forest type, and $\hat{v}_{tsd}$ is the volume per acre for combinations of species groups and diameter classes.

If trees in an original forested condition occurred on a plot that was primarily nonforested, using the one condition per plot method removed these trees from the calculations.

Results

Land Area

Because the total forested area is determined by ground truthed photo interpretation points, using rule 2 does not affect the estimates for the total forested area because rule 2 does not change any ground truth calls from forested to nonforested or from nonforested to forested. However, using rule 1 does change some of the ground truth calls for phase 1 photo interpretation points. Using rule 1 increased the estimated total forested area for the State by 0.4 percent and decreased the estimated variance by 0.4 percent. The table cells with the largest area, primarily marginal totals for either forest types or owner groups, had small percent changes, ranging from –6.8 percent to + 2.5 percent. The smaller combinations of forest type and ownership group had larger percent changes, ranging from –100.0 percent to +18.4 percent. The effects of rule 2 on the percent changes followed the same pattern as the effects of rule 1.

Volume

Using rule 1 decreased the estimate for total volume by 3.9 percent and decreased the variance by 15.6 percent. As with the effects on area, the table cells with the greatest amount of volume, primarily the marginal totals for species or diameter classes, show small percent changes, from –6.6 percent to –1.7 percent. The table cells with small amount of volume showed large percent changes, ranging from –100.0 percent to +26.7 percent.

Using rule 2 caused no changes in the estimated volumes from the current method but did change the variance estimates. Using rule 2 decreased the variance of the estimate for total volume by 12.9 percent.

Discussion

Using rule 2 did change the individual cell estimates for area, but not for volume. The reason the volume estimates do not change can be found in equation 7. A close look at the term $\hat{R}_t * \hat{v}_{tsd}$ reveals the cause. In equation 9, the numerator of the first quotient is the area in forest type t summed over all forested conditions, and the denominator of the second quotient is the area in forest type t summed over all forest conditions in forest type t.

$$
\sum_{t=1}^{T} \hat{R}_t * \hat{v}_{tsd} = \sum_{t=1}^{T} \frac{\sum_{i=1}^{n_f} A_{ti}}{\sum_{i=1}^{n_f} A_i} * \frac{\sum_{i=1}^{n_t} v_{tsdi}}{\sum_{i=1}^{n_t} A_{ti}} \tag{9}
$$

Since the area in forest type t on plots without any forest type t is 0, these two numbers are equal and will cancel out. Since this is the only condition variable in this equation, the estimates of volume will not depend on how the forest types are determined.

The 12.9-percent decrease in the variance of the estimated total volume for rule 2 is the result of merging forested conditions. When conditions are merged, the resulting volume per acre will be a weighted average of the volume per acre of the conditions that are merged. This weighted average will remove some of the variation in the volume per acre values.

Using rule 1 did affect the estimates of both State totals: an increase in total forest area of less than 0.2 percent and a decrease in total volume of 3.9 percent. When rule 1 is used, some forested and nonforested conditions will be relabeled. The increase in the estimated total forested area suggests that more nonforested area was relabeled as forested than forested area was relabeled as nonforested. When forest is relabeled as nonforest, the trees get removed from the total volume estimate. When nonforest is relabeled as forest, they will contribute to the area estimates, but they don't have any trees to add to the volume estimates, causing a decrease in the volume per acre values for that plot. This decreases the estimated total volume. Changes in the estimates for the individual cells on the reporting tables were much larger, ranging from −100 percent to +30 percent.

Arner (1998) ran a similar study with data from Maine on a different fixed-plot design. His design was a single 1/5-acre plot instead of four 1/24-acre plots. His totals also changed very little and showed larger percent changes in the smaller categories. His relative changes were smaller due in part to the fact that he reported just marginal totals, and in part to the fact the he had more plots. Even though Maine is a smaller state, Arner had more plots because he used the data from a statewide survey, while our study used only one of the panels from South Carolina. Using the rest of the panels from South Carolina would probably lead to smaller percent changes.

In South Carolina, 30 to 35 percent of the plots had more than one condition, but only about 8 percent of the subplots had condition boundaries. Of that 8 percent, about two-thirds were forest-nonforest boundaries. This suggests that using subplot lumping of conditions using either rule 1 or rule 2 methods will also lead to smaller percent changes. Unfortunately, one piece of information required to run this analysis on the subplot level is not in the database for South Carolina's first three panels.

Conclusions

The effect of rule 1 (one condition per plot) on the total area estimate is primarily due to random error. However, rule 1 did bias the estimate for total volume and therefore is unacceptable as an alternative to the current methods. Rule 2 (one forested condition per plot) did not change the estimates for either total forested area or total volume and decreased the variance for the total volume. Rule 2 also did not change the estimates for the volume by species and diameter class combinations. Therefore, rule 2 appears to be an acceptable alternative to the current method.

Recommendations

To determine if the current procedures ought to be changed, we suggest the following steps:

1. To more accurately determine the size of the effects of using rule 2, a similar study should be completed using five sequential panels that include the data required to analyze the effects on the subplot-condition instead of the plot-condition.
2. The results of that study should be discussed with users to determine the effect of the changes in the tables, such as the area by owner group-forest combination table, on the users.
3. The effect of using rule 2 on growth projections should be studied, because some of the condition differences are important to growth projections. For instance, if a 5-inch-diameter loblolly pine stand is growing next to a 9-inch-diameter loblolly pine stand, you would assume that the trees would grow differently in these separate stands than if they were in a single, two-aged stand.

These three steps should allow FIA to make an informed decision about changing to just one forested condition per subplot.

Literature Cited

Arner, Stanford L. 1998. Comparison of a fully mapped plot design to three alternative designs for volume and area estimates using Maine inventory data. Gen. Tech. Rep. NE-243. Upper Darby, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 7 p.

Cochran, William G. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons. 427 p.

FIA Stat Band (Scott, Charles T.; Bechtold, William A.; Reams, Gregory A.; Smith, William D.; Hansen, Mark H.; Moisen, Gretchen G.). In review. Forest inventory and analysis: national sample design and estimation procedures. Bechtold, William A., and Patterson, Paul L., eds. At http://srsfia1.fia.srs.fe.fed.us/statistics_band/stat_documents.htm.

Goodman, Leo A. 1960. On the exact variance of products. Journal of the American Statistical Association. 55: 708–713.

McCollum, Joseph. 2002. Standard errors in forest area. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; Moser, John W., eds. Proceedings, 3d annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 5–10.

Reams, Gregory. 2000. SAFIS area estimation techniques. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C., eds. Proceedings, 1st annual forest inventory and analysis symposium; 1999 November 2–3; San Antonio, TX. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 32–36.

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

Thompson, Michael. 1998. Forest statistics for Georgia, 1997. Resour. Bull. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 92 p.

Zarnoch, Stan; Bechtold, William. 2000. Estimating mapped-plot forest attributes with ratios of means. Canadian Journal of Forest Research. 30: 688–697.