ANNUAL FOREST INVENTORY ESTIMATES BASED ON THE MOVING AVERAGE

Francis A. Roesch, James R. Steinman, and Michael T. Thompson[1]

ABSTRACT.—Three interpretations of the simple moving average estimator, as applied to the USDA Forest Service's annual forest inventory design, are presented. A corresponding approach to composite estimation over arbitrarily defined land areas and time intervals is given for each interpretation, under the assumption that the investigator is armed with only the spatial/temporal matrix of moving average estimates. The advantages and practical limitations of each interpretation are discussed.

After a long history of conducting periodic inventories on a State-by-State basis, the USDA Forest Service's Forest Inventory and Analysis (FIA) program has initiated an annual forest inventory sampling design in which all forest land is sampled each year. The FIA units at the six Forest Service Research Stations are at various stages of implementing this common sampling design. The program at the Southern Research Station (SRS) has been known as the Southern Annual Forest Inventory System (SAFIS), e.g., see Roesch and Reams (1999). SAFIS was introduced to improve estimation of both the current resource inventory and changes in the resource. Within FIA, it is assumed that users will desire the ability to form estimates over arbitrary land areas and time intervals. It is also assumed that, because data are collected every year, users will expect to be able to obtain new information on an annual basis. However, the primary charge to FIA is to report on the state of the forest resource every 5 years using the data collected over the past 5 years. The design and intensity of the sample are intended to provide the ability to form estimates using 5 years' worth of data with the same degree of confidence as that available from the previous periodic surveys. FIA recognizes that a complete set of data will be available annually after the first 5 years of data have been collected but does not expect to have the resources necessary to publish complete reports annually. One proposal has been to make intermediate estimates based on the moving average estimator available to a wide audience over the Internet. This would probably lead to questions as to how the moving average estimator should be interpreted, and how it should be used to form annual estimates and estimates of change over various time intervals. Below we present three simple interpretations of the moving average estimator along with their corresponding methods of making estimates over arbitrarily defined land areas. We also briefly mention some drawbacks to these approaches.

The plot arrangement for the annual FIA 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, e.g., Overton and others (1990) and White and others (1992). The sample plots are located in a systematic triangular grid with five interpenetrating panels. One panel per year is measured for 5 consecutive years. Every 5 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. Figure 1 depicts the data availability from the panels for the first 10 years of measurement.

In the sequel, we assume that the user is interested in a specific classification of a particular area and time interval. Specifically, the classification must uniquely partition the area into mutually exclusive, all-inclusive classes. With respect to the sample, this requires the user to define the classification in such a way that a variable or a set of variables available in the data can be used to partition the sample into classes at discrete points in time. We show how the user could form estimates for each class within the defined area and time interval.

[1] Respectively, Mathematical Statistician and Research Forester, Southern Research Station, USDA Forest Service, 160A Zillicoa Street, P.O. Box 2750, Asheville, NC 28802-2750; and Forester, USDA Forest Service, 200 W. T. Weaver Blvd., Asheville, NC 28804.

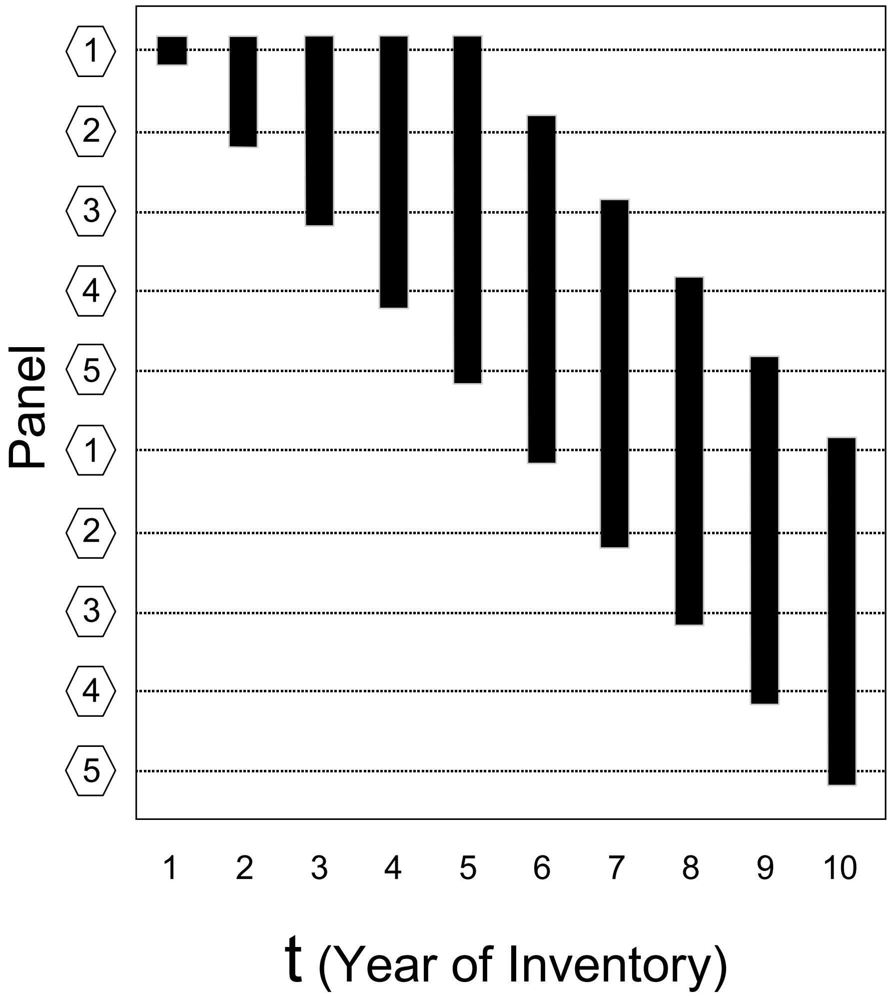

Figure 1.—*Panel data availability from the five-panel annual FIA design. The panel enumeration represents successive years of measurement.*

For example, one of the measured variables for each plot is the proportion of plot area in each forest condition class. 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). The user might be interested in forming estimates for each forest condition class within a particular area and time period. Explicitly, the user has no interest in any area that is not defined as forest. Whatever classification is chosen, we recognize that it could be one that continuously changes through time, such as the above example that puts all nonforested areas into one class but separates forested areas into unique condition classes.

For practical purposes we need to fix the classification over discrete time intervals. In this paper, we will assume that the minimum usable interval is 1 year. This might be perceived as somewhat of a break with tradition since FIA has historically treated the classification as fixed over the entire measurement interval, which often consisted of several years. Since it takes 5 years to measure all of the plots in the sample, the measurement interval is 5 years, but a new measurement interval is defined every year after the fifth year. These observations present opportunity in the guise of a dilemma. The dilemma exists because the longer measurement interval allows more time for potentially significant changes to occur in the underlying population classification. If we treat the classification as fixed over the measurement interval, the perception of the sample with respect to the classification will be simplified. However, there will be no way to estimate trends such as changes in the proportion of area in each class. If part of the classification depends upon information gathered at a fixed point in time, such as a pre-cycle aerial photo interpretation, the expected value of the "true class" given the assumed or fixed class would show an increase in bias as time moves away from the time of the classification.

If we treat the classes as fixed yearly, the perception of the sample is complicated by the acknowledgement that each panel is potentially sampling a different underlying classification. The result is that the minimum area over which we make estimates by class may have to be somewhat larger to ensure adequate sample sizes. The opportunity presented by the dilemma is the ability to account for changes occurring in the underlying classification during the measurement interval, at least over sufficiently large areas. At the same time, within class variance will be minimized because the "true" class will usually equal the assumed class.

NOTATION

Let

V_{ikt} = the value observed at plot i ($i = 1,...,I$), in class k, at time t

A_{ikt} = the area in acres of plot i sampled in class k at time t,

$$C_{ikt} = \begin{cases} 1 & \text{If class } k \text{ occurs at plot } i \ (i = 1,...,I) \text{ and time } t, \\ 0 & \text{Otherwise} \end{cases}$$

n_t = the number of plots within the arbitrarily selected boundary at time t

n_c = the number of classes

$A_{i\bullet t} = \sum_{k=1}^{n_c} A_{ikt}$ = the area in acres sampled at plot i ($i = 1,...,I$), and time t,

$$n_{\bullet kt} = \sum_{i=1}^{n_t} C_{ikt}$$

$$A_{\bullet kt} = \sum_{i=1}^{n_t} A_{ikt}$$

$$\overline{V}_{ikt} = \begin{cases} V_{ikt} / A_{ikt} & \text{If class } k \text{ occurs at plot } i \text{ and time } t, \\ 0 & \text{otherwise} \end{cases}$$

$$\overline{V}_{\bullet kt} = \sum_{i=1}^{n_t} A_{ikt}\overline{V}_{ikt} \Big/ A_{\bullet kt} = \sum_{i=1}^{n_{kt}} V_{ikt} \Big/ \sum_{i=1}^{n_{kt}} A_{ikt}$$, for sampled class k.

INTERPRETATIONS OF THE MOVING AVERAGE

Suppose that the user's focus is on estimation of per acre value (V) and change in that value for each class k. However, for whatever reason, the original annual data are not available to the user. What is available to the user is the overall mean for the five-panel series, which is one estimator of per acre value (V) for class k

$$\hat{V}_{\bullet k \bullet} = \frac{1}{5}\sum_{t=h}^{h+4} \overline{V}_{\bullet kt}$$

Each year, new data are available for one of the five panels. The previous measurements for that panel are deleted from the estimator and the new measurements are incorporated (h is increased to $h+1$); therefore, the estimates are formed as a 5-year window moves through time, hence the term moving average.

Specifically, what does the moving average estimate with respect to time? Well, that depends upon the individual's viewpoint. We'll give three potential interpretations, each of which leads to a method of estimating change. Each of these methods could be used even if the annual data have not been provided and one has only the series of moving averages available.

One could assume that the moving average is providing an estimate of the center of the time interval (viewpoint 1). Alternatively, one could assume that the moving average is providing an estimate for the entire time interval (viewpoint 2). A third viewpoint is that the moving average is the mean of an unknown time-dependant linear combination, such as a simple linear trend (viewpoint 3).

Viewpoint 1:

Considering viewpoint 1, illustrated in figure 2, the moving average is an unbiased estimator for the center year if there is either no trend in the variable of interest or if the trend is linear. We'll represent the moving average midpoint estimator as *MM* subscripted by the middle year, i.e., the mean of values for years 1995 through 1999 would be MM_{1997} so an estimator of change between two arbitrary years, say year 1 and year 2 would simply be

$$CM_{Y_1Y_2} = MM_{Y_2} - MM_{Y_1}$$

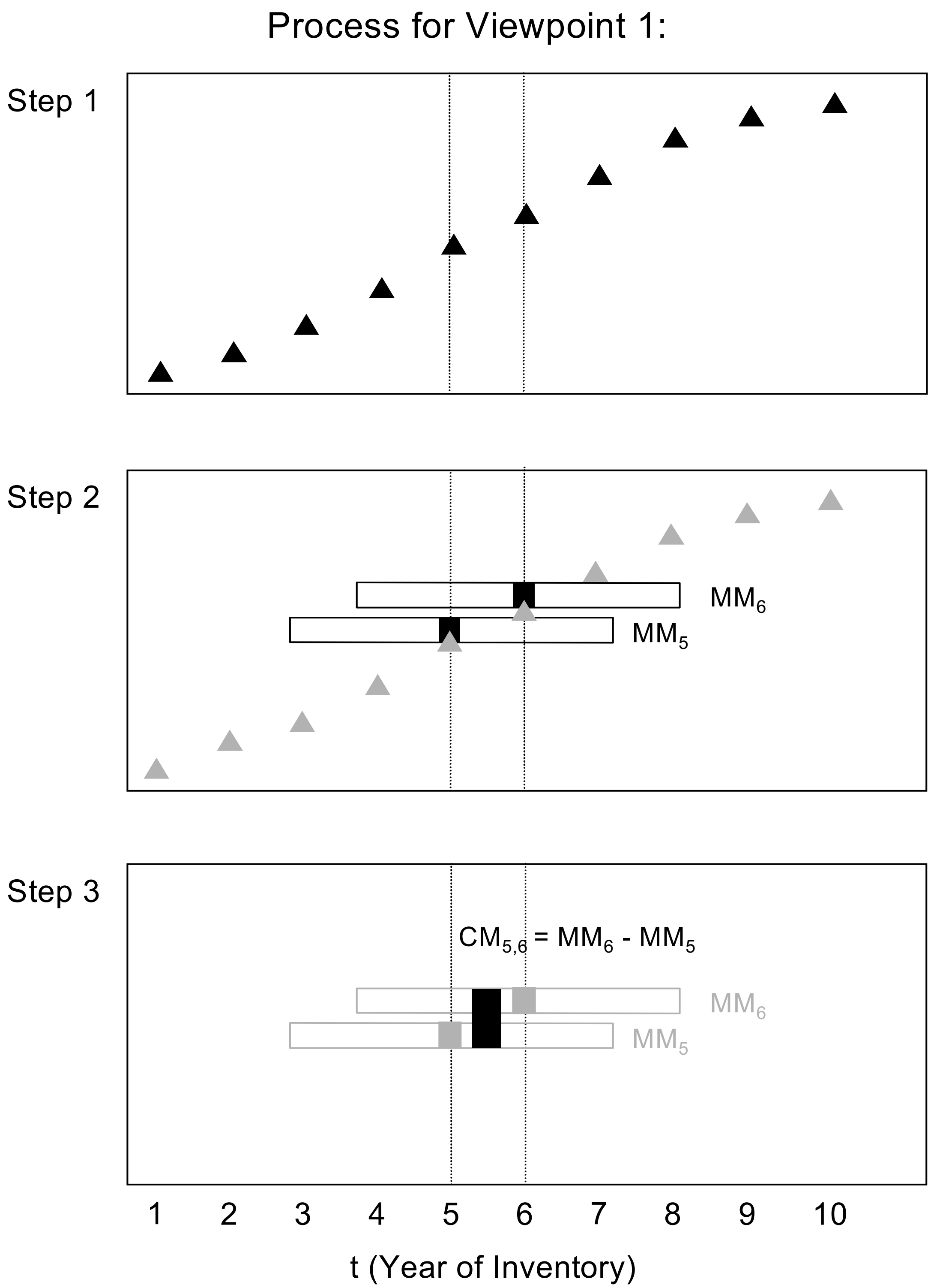

Figure 2.—*Step 1 shows hypothetical means from the original data by year. Step 2 shows the years over which two moving averages were made (empty rectangle), while the dark squares represent the interpretation of the moving average as an estimator of the midpoint of the time interval. Step 3 depicts change as the difference between the two midpoint estimators.*

The major drawback to MM_Y as an estimator is that it cannot be shown to be unbiased in the presence of any trend other than a simple linear trend. However, the change estimator $CM_{Y_1Y_2}$ will be unbiased whenever the bias in MM_Y is zero or constant. A disadvantage to $CM_{Y_1Y_2}$ is that the first year of interest would need to be after the second year of data collection and at least 2 years of data must be collected after the final year of interest.

Viewpoint 2:

Given viewpoint 2, illustrated in figure 3, the moving average is an unbiased estimator for the interval mean over a restricted set of trend assumptions. We'll represent the moving average interval estimator as *MI* subscripted by the first and last year of the interval, i.e., the mean of values for years 1995 through 1999 would be $MI_{1995,1999}$. To form annual estimates, we note that eventually, for a particular year of interest (y), there would be five interval estimates that cover the year, $MI_{y-4,y}$, $MI_{y-3,y+1}$, $MI_{y-2,y+2}$, $MI_{y-1,y+3}$, and $MI_{y,y+4}$. An estimator for year *y* could be formed by taking the mean of these five interval estimates

$$MA_y = \frac{1}{5}\sum_{i=4}^{0} MI_{y-i;y-4+i}$$

The change estimate is found through subtraction

$$CA_{Y_1Y_2} = MA_{Y_2} - MA_{Y_1}$$

The largest drawback to using $CA_{Y_1Y_2}$ is that the first year of interest would need to be greater than the fourth year of data collection and at least 4 years of data must be collected after the final year of interest. This is not an absolute restriction, since adaptation of the estimator to fewer than five available intervals is trivial.

Viewpoint 3:

With viewpoint 3, one has a route to annual estimates as soon as the second average has been obtained using a difference estimator. We give an illustration of this viewpoint in figure 4. We'll represent the moving average linear estimator as *ML* subscripted by the first and last year of the interval, i.e., the mean of values for years 1995 through 1999 would be $ML_{1995,1999}$. To form annual estimates, we note that the difference, $d_t = ML_{t+1,t+5} - ML_{t,t+4}$, under a simple linear model leads to an immediate series of annual estimates:

$$ML_{t+1,t+5} = ML_{t,t+4} + d_t$$

$$\sum_{i=t+1}^{t+5}\frac{AL_i}{5} = \sum_{i=t}^{t+4}\frac{AL_i}{5} + d_t$$

$$\left(\sum_{i=t+1}^{t+4}\frac{AL_i}{5}\right) + \frac{AL_{t+5}}{5} = \frac{AL_t}{5} + \left(\sum_{i=t+1}^{t+4}\frac{AL_i}{5}\right) + d_t$$

$$\frac{AL_{t+5}}{5} = \frac{AL_t}{5} + d_t$$

$$AL_{t+5} = AL_t + 5d_t \therefore$$

$$AL_{t+4} = AL_t + 4d_t$$

$$AL_{t+3} = AL_t + 3d_t$$

$$AL_{t+2} = AL_t + 2d_t$$

$$AL_{t+1} = AL_t + 1d_t$$

$$AL_t = AL_t + 0d_t$$

To start the recursive series, we need the annual estimator AL_t in terms of $ML_{t,t+4}$

$$ML_{t,t+4} = ML_{t+1,t+5} - d_t$$

$$ML_{t,t+4} = \frac{AL_{t+1} + AL_{t+2} + AL_{t+3} + AL_{t+4} + AL_{t+5}}{5} - d_t$$

$$ML_{t,t+4} = \frac{AL_t + d_t + AL_t + 2d_t + AL_t + 3d_t + AL_t + 4d_t + AL_t + 5d_t}{5} - d_t$$

$$ML_{t,t+4} = AL_t + 2d_t$$

$$AL_t = ML_{t,t+4} - 2d_t$$

A linear predictor is available for any successive year *h*, $AL_{t+h} = AL_t + hd_t$. Therefore, the change estimator between any two years *t* and *t+h*

$$CL_{t+h,h} = AL_{t+h} - AL_t = hd_t \ .$$

The development above gives us a prior prediction for year *t+6*

$$AL_{t+6} = AL_t + 6d_t$$

Another estimate for time *t+6* can be found once the estimator $ML_{t+2,t+6}$ is available by reapplying the above logic

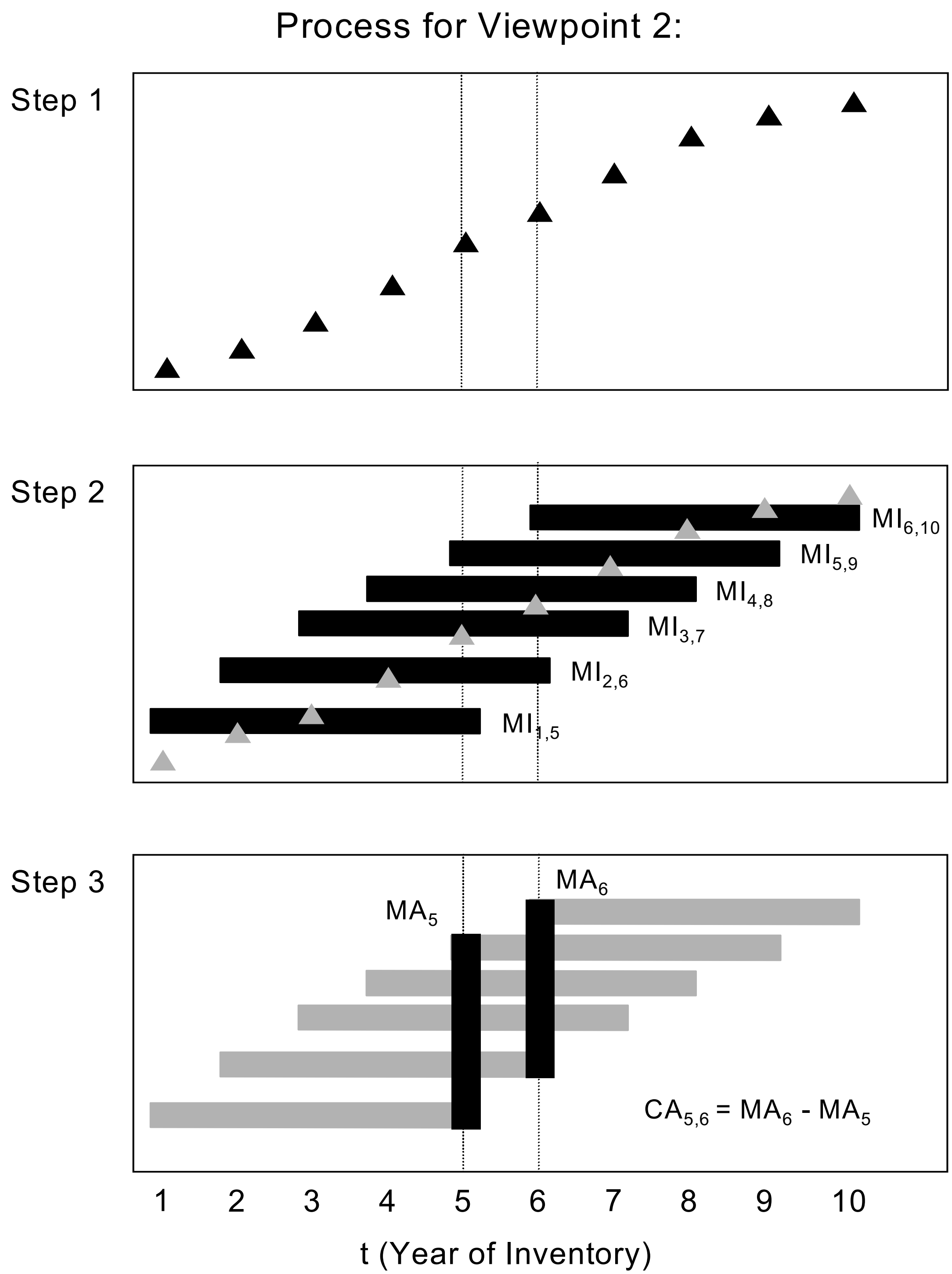

Figure 3.—*As in figure 2, step 1 shows hypothetical means from the original data by year. Step 2 shows the years over which six moving averages were made (filled rectangles), with the interpretation of the moving average as an estimator of the interval. Step 3 depicts the means of all five intervals covering each of the same 2 years of interest shown in figure 2. Change is given as the difference between the two mean of intervals estimators.*

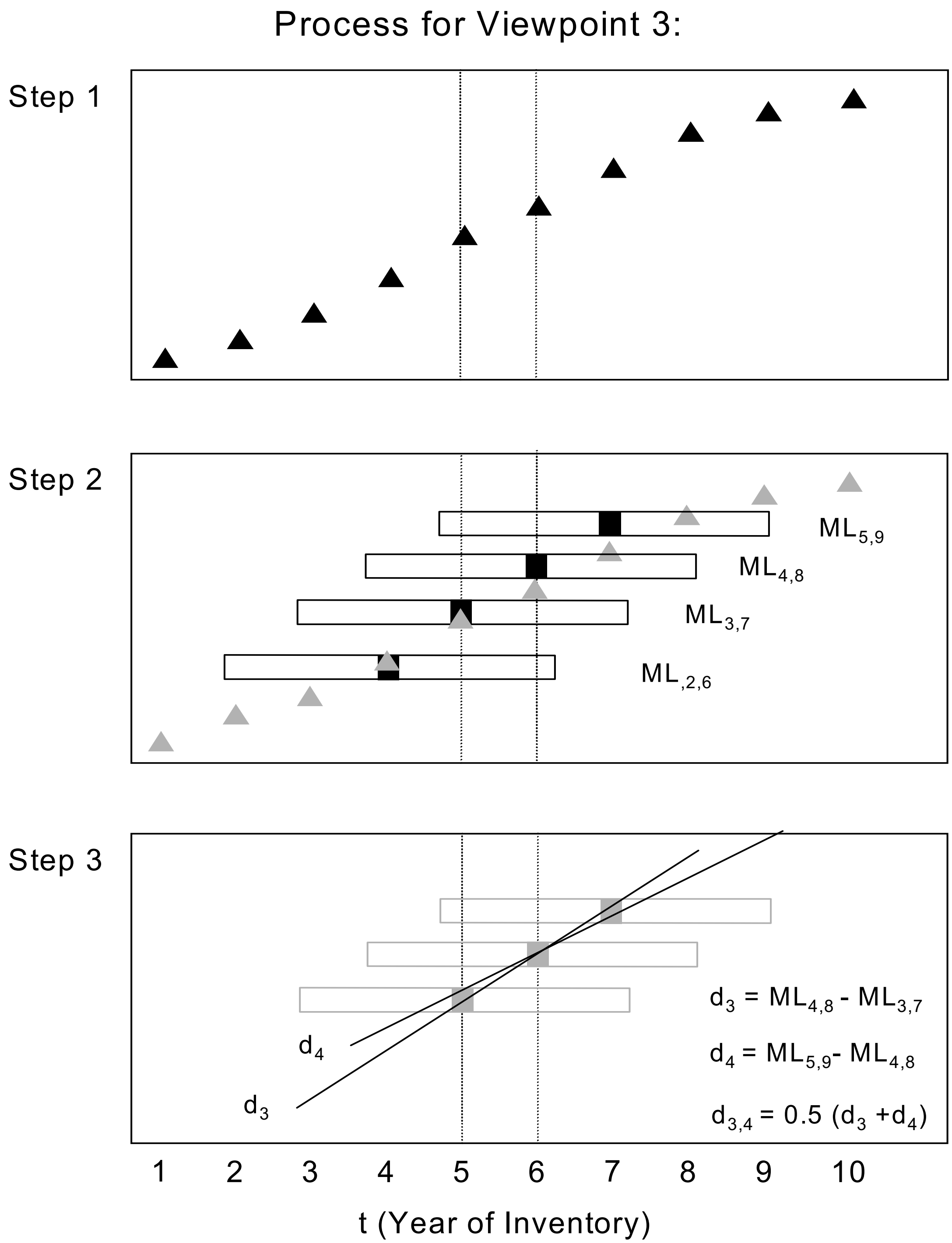

Figure 4.—*As in figures 2 and 3, step 1 shows hypothetical means from the original data by year. Step 2 shows the years over which four moving averages were made (empty rectangles), with a point representing the average arbitrarily shown at the center of the interval (solid square). The interpretation is that each moving average is the result of an unknown linear combination, which depends upon year of measurement. Step 3 depicts two estimates of the trend found by taking the difference between successive moving averages. Finally, a mean trend is shown as the average of two successive trends.*

$$ML_{t+2,t+6} = ML_{t+1,t+5} + d_{t+1}$$

$$\sum_{i=t+2}^{t+6} \frac{AL_i}{5} = \sum_{i=t+1}^{t+5} \frac{AL_i}{5} + d_{t+1}$$

$$\left(\sum_{i=t+2}^{t+5} \frac{AL_i}{5}\right) + \frac{AL_{t+6}}{5} = \frac{AL_{t+1}}{5} + \left(\sum_{i=t+2}^{t+5} \frac{AL_i}{5}\right) + d_{t+1}$$

$$\frac{AL_{t+6}}{5} = \frac{AL_{t+1}}{5} + d_{t+1}$$

$$AL_{t+6} = AL_{t+1} + 5d_{t+1}$$

$$AL_{t+6} = AL_{t+1} + 5d_{t+1} \;\therefore$$

$$AL_{t+5} = AL_{t+1} + 4d_{t+1}$$

$$AL_{t+4} = AL_{t+1} + 3d_{t+1}$$

$$AL_{t+3} = AL_{t+1} + 2d_{t+1}$$

$$AL_{t+2} = AL_{t+1} + 1d_{t+1}$$

$$AL_{t+1} = AL_{t+1} + 0d_{t+1}$$

Noting that similar to the above, we can find the start point

$$AL_{t+1} = ML_{t+1,t+5} - 2d_{t+1}$$

Then

$$AL_{t+1} = ML_{t+1,t+5} - 2d_{t+1}$$

$$AL_{t+2} = ML_{t+1,t+5} - d_{t+1}$$

$$AL_{t+3} = ML_{t+1,t+5}$$

$$AL_{t+4} = ML_{t+1,t+5} + d_{t+1}$$

$$AL_{t+5} = ML_{t+1,t+5} + 2d_{t+1}$$

$$AL_{t+6} = ML_{t+1,t+5} + 3d_{t+1}$$

Setting the two estimates for AL_{t+6} equal

$$AL_{t+6} \Rightarrow ML_{t,t+4} + 4d_t \equiv ML_{t+1,t+5} + 3d_{t+1}$$

$$\Rightarrow ML_{t+1,t+5} + 3d_t \equiv ML_{t+1,t+5} + 3d_{t+1}$$

$$\Rightarrow d_t \equiv d_{t+1}$$

If $d_t \neq d_{t+1}$, an adjustment to the linear model is suggested. Retaining the linearity assumption, we could use the mean slope

$\overline{d}_{t,t+1} = .5(d_t + d_{t+1})$, and the mean midpoint

$$\overline{ML}_{t+1,t+5} = \frac{1}{3}\left(ML_{t,t+4} + ML_{t+1,t+5} + ML_{t+2,t+6}\right)$$

to obtain new estimates for each year

$$\overline{AL}_t^{t,t+1} = \overline{ML}_{t+1,t+5} - 3\overline{d}_{t,t+1}$$

$$\overline{AL}_{t+1}^{t,t+1} = \overline{ML}_{t+1,t+5} - 2\overline{d}_{t,t+1}$$

$$\overline{AL}_{t+2}^{t,t+1} = \overline{ML}_{t+1,t+5} - \overline{d}_{t,t+1}$$

$$\overline{AL}_{t+3}^{t,t+1} = \overline{ML}_{t+1,t+5}$$

$$\overline{AL}_{t+4}^{t,t+1} = \overline{ML}_{t+1,t+5} + \overline{d}_{t,t+1}$$

$$\overline{AL}_{t51}^{t,t+1} = \overline{ML}_{t+1,t+5} + 2\overline{d}_{t,t+1}$$

$$\overline{AL}_{t+6}^{t,t+1} = \overline{ML}_{t+1,t+5} + 3\overline{d}_{t,t+1}$$

This crude but effective approach could be reapplied as each successive year's average became available.

Again, the change estimate is found through subtraction

$$CL_{Y_1,Y_2} = \overline{AL}_{Y_2}^{t,t+k} - \overline{AL}_{Y_1}^{t,t+k} \quad .$$

The advantage of viewpoint 3 over viewpoint 2 is that the first year of interest could be the first year that data are collected, as long as 6 years of data have been collected, and the final year of interest could be the final year of data collection. The major disadvantage is the potential of a poor fit to other than simple linear trends.

NUMERICAL EXAMPLE

In figures 5 and 6, we present a numerical example that illustrates estimates of softwood volume per acre in planted stands. Although the data are contrived, the range of observations could realistically be expected in an area such as southeast Georgia. The values in figure 5 are estimates from successive 5-year intervals, analogous to the horizontal bars in figures 2 through 4. Figure 6 illustrates the mean trend of viewpoint 3 for the data in figure 5 compared to the individual annual estimates.

CONCLUSION

The development above is intended to shed some light on the potential effects of providing moving average estimates. On the one hand, we have a sample design spread uniformly through time and space, and this design provides adequate data to make estimates over domains of interest every 5 years. On the other hand, we expect that many users will be interested in uniquely

Year	Softwood Volume	5-Year Moving Average	Viewpoint 1- Moving average as a Midpoint Estimator		Change Estimator		Viewpoint 2- Moving average as an Interval Estimator		Mean of Covering Intervals	
1	1570.0									
2	1459.6									
3	1501.2		MM_3	1477.0			$MI_{1,5}$	1477.0		
4	1350.6		MM_4	1443.2	$CM_{3,4}$	-33.8	$MI_{2,6}$	1443.2		
5	1503.5	1477.0	MM_5	1431.2	$CM_{4,5}$	-12.1	$MI_{3,7}$	1431.2	MA_5	1441.0
6	1401.2	1443.2	MM_6	1423.2	$CM_{5,6}$	-8.0	$MI_{4,8}$	1423.2	MA_6	1430.3
7	1399.3	1431.2	MM_7	1430.2	$CM_{6,7}$	7.0	$MI_{5,9}$	1430.2		
8	1461.4	1423.2	MM_8	1423.6	$CM_{7,8}$	-6.6	$MI_{6,10}$	1423.6		
9	1385.6	1430.2								
10	1470.5	1423.6								

Year	Softwood Volume	5-Year Moving Average	Viewpoint 3- Moving average as a General Linear combination				First Two Available		Second and Third		Combined	
1	1570.0						AL_1	1544.5	Predict AL^2_1	1479.4	$AL^{1,2}_1$	1519.18
2	1459.6						AL_2	1510.7	AL^2_2	1467.3	$AL^{1,2}_2$	1496.27
3	1501.2		$ML_{1,5}$	1477.0			AL_3	1477.0	AL^2_3	1455.3	$AL^{1,2}_3$	1473.36
4	1350.6		$ML_{2,6}$	1443.2	d_1	-33.8	AL_4	1443.2	AL^2_4	1443.2	$AL^{1,2}_4$	1450.45
5	1503.5	1477.0	$ML_{3,7}$	1431.2	d_2	-12.1	AL_5	1409.5	AL^2_5	1431.2	$AL^{1,2}_5$	1427.54
6	1401.2	1443.2	$ML_{4,8}$	1423.2	d_3	-8.0	AL_6	1375.7	AL^2_6	1419.1	$AL^{1,2}_6$	1404.63
7	1399.3	1431.2	$ML_{5,9}$	1430.2	d_4	7.0	Predict AL_7	1341.9	AL^2_7	1407.0	$AL^{1,2}_7$	1381.72
8	1461.4	1423.2	$ML_{6,10}$	1423.6	d_5	-6.6						
9	1385.6	1430.2							$\sim d_{1,2}$	-22.91		
10	1470.5	1423.6							$\sim ML_{2,6}$	1450.45		

Figure 5.—*Estimates formed from each of the three viewpoints.*

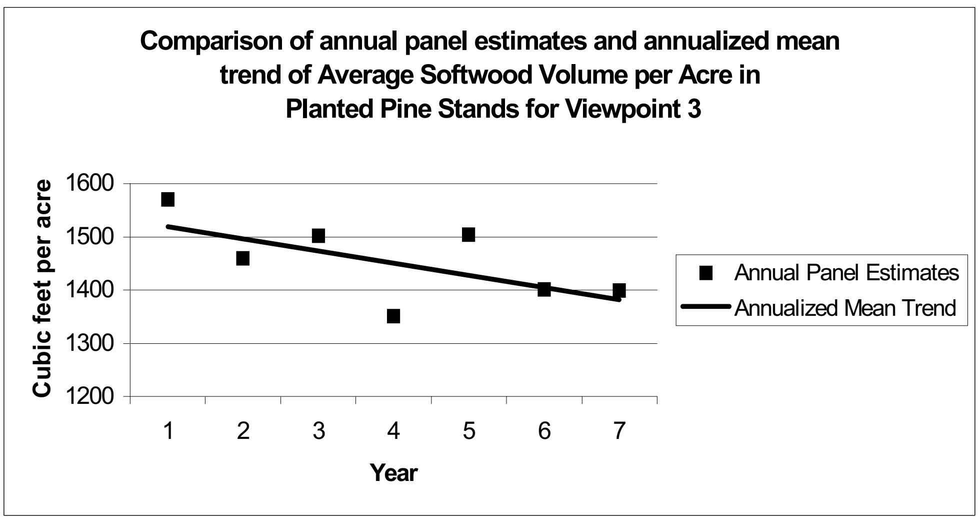

Figure 6.—*The mean trend plotted for the data in figure 5.*

defined subareas and time intervals. Specifically, many users will expect annual estimates. For users with the wherewithal to analyze the raw data, we could simply provide the raw data. Users could then choose to model the time trend if they so desire. 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. Roesch (2001) discusses this and other approaches in the simpler context of a single panel series. However, many users would rather obtain the information in some reduced form. Because of the wide variety of variables that we report on, it would be difficult to identify a single temporal model that would be applicable to all variables of interest. This has resulted in an internal agreement to use the moving average when a specific temporal model has not been identified.

In this paper, we discussed three methods for interpreting the moving average over arbitrary intervals of space and time from the annual FIA design. These are not the estimators that FIA will use in its processing system for standard estimates. Over the next few years, FIA will be evaluating models of varying complexity for all variables of interest in an effort to determine the most efficient approach for each variable.

LITERATURE CITED

Anonymous. 1998. Field instructions for Southern Forest Inventory, version #2. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 171 p.

Overton, W.S.; White, D.; Stevens, D.L. 1990. Design report for EMAP (Environmental Monitoring and Assessment Program). EPA/600/3-91/053. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development.

Roesch, F.A. 2001. Mixed estimation for a forest survey sampling design. In: Proceedings, Section on statistics and the environment. Baltimore, MD: Joint Statistical Meetings. August 8-12, 1999.

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

Van Deusen, P.C. In review. Modeling trends with annual survey data.

White, D.; Kimberling, J.; Overton, S.W. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartography and Geographic Information Systems. 19: 5-22.