Comparison of Estimators for Rolling Samples Using Forest Inventory and Analysis Data

Devin S. Johnson, Michael S. Williams, and Raymond L. Czaplewski

ABSTRACT. The performance of three classes of weighted average estimators is studied for an annual inventory design similar to the Forest Inventory and Analysis program of the United States. The first class is based on an ARIMA(0,1,1) time series model. The equal weight, simple moving average is a member of this class. The second class is based on an ARIMA(0,2,2) time series model. The final class is based on a locally weighted least-squares regression prediction. The estimator properties were tested using a simulation population created from Forest Inventory and Analysis (FIA) data from northeastern Minnesota. Estimates of total volume per acre, on-growth volume per acre, mortality volume per acre, proportion of sawtimber acreage, proportion of poletimber acreage, and proportion of sapling acreage were calculated using several weighted average estimators in each year. These were compared to the simulation population, for which the true values are known, and an unbiased yearly estimator. When computing estimates, the ARIMA(0,1,1) based estimators produced the lowest root mean squared error of each of the three classes. However, in a few years the bias for some variables was high. The maximum percent increase between the estimator with the lowest root mean squared error and the simple moving average was 7.31%. Of all the estimators, the simple moving average performed well in terms of mean square error in virtually every situation. It tended to be best among the estimators tested if spatial variation was large and change was relatively small. It was not consistently best in terms of mean square error in the presence of moderate change and large spatial variation. For. Sci. 49(1):50–63.

Key Words: Rolling samples, panel survey, annual inventory, FIA, weighted average.

The USDA Forest Service Forest Inventory and Analysis (FIA) program provides information on the status and trends of the nation's forest resources. Historically, surveys were conducted periodically on a state-by-state basis with the time between inventories ranging from 6 to 18 yr depending on the region of the country (Gillespie 1999). In recent years, however, consumers of this information have been concerned that FIA is not adequately meeting their needs for current data (Van Deusen et al. 1999). Recent legislation has directed FIA to switch to an annual, or panel survey, where a proportion of each state's inventory is completed every year.

Under periodic surveys, population statistics were estimated independently from cycle to cycle. Estimates from any given cycle possessed a high level of precision. Since cycle lengths were many years apart, previous estimates contained little information about current conditions. In an annualized survey, however, previous data are only a few years old, so they may contain a significant amount of information about conditions at the time of the present survey. Yearly estimates will, however, have less precision than the periodic surveys

Devin S. Johnson, Department of Statistics, Colorado State University, Fort Collins, Colorado 80523—Phone: (970) 491-6014; Fax: (970)491-7895; E-Mail: johnson@stat.colostate.edu. Michael S. Williams and Raymond L. Czaplewski, USDA Forest Service, Rocky Mountain Research Station, 2150 Centre Ave., Bldg A, Suite 361, Fort Collins, Colorado 80526.

Acknowledgments: The authors are grateful to the associate editor, as well as two anonymous reviewers, whose comments contributed to the paper. This research was partially supported by NSF grant DUE-0094514.

Manuscript received April 4, 2001, accepted October 30, 2001.

due simply to smaller sample measured each year. With this in mind, it has often been suggested that estimates for current conditions should take advantage of previous data (Czaplewski 1995, Gillespie 1999, McRoberts 1999, McRoberts and Hansen 1999, Reams and Van Deusen 1999, Roesch and Reams 1999, Van Deusen et al. 1999).

There have been several different estimation strategies proposed that would take advantage of previous data (Czaplewski 1995 and Reams and Van Deusen 1999). One estimator of the current population statistic, θ_t, which appears often in the literature is the rolling average (*RA*) given by

$$\hat{\theta}_{RA,t} = \sum_{h=1}^{T} w_{t-T+h} \hat{\theta}_{t-T+h}$$

where $\hat{\theta}_{t-T+h}$, $h = 1,\ldots,T$ is an unbiased estimator of θ_{t-T+h}, the population characteristic at time $t - T + h$, and $\{w_{t-T+h}\}$ is a set of constants which sum to 1 (McRoberts 1999, Roesch and Reams 1999). The constant, T, is equal to the number of years until the entire inventory is completed. If 20% of the plots selected are sampled each year then $T = 5$.

The rolling estimator has many advantages. Under the randomization paradigm described by Eriksson (1995, 2001a), $\hat{\theta}_{RA,t}$ is easy to calculate from inventory data and has a variance estimator that is relatively easy to calculate from data as well. The RA estimator will have a lower annual variation than any one of the yearly estimators. This fact may appeal to consumers of FIA data, who may distrust estimates that vary too much from year to year (Gillespie 1999).

The major problem with the *RA* estimator is that it is biased as an estimator of the current year's population parameter. The rationale for the rolling average estimator is that by using past data you are "borrowing strength," in terms of sample size, from previous years. While this creates a "lag" bias when estimating current conditions, it is more than compensated for by a reduction in variance. This unequal trade between variance and bias leads to a lower mean squared error for the rolling average estimator.

Another problem is selection of the weights. Roesch and Reams (1999) suggest that equal weights could be used for each year, although they also state that equal weighting might mask time trends. However, the equal weighting estimator can be thought of as an unbiased estimator for the population parameter at some time approximately in the middle of the rotation cycle, so, if the time trends are long enough in duration, the equal weighting estimator will react to the trends. Eriksson (2001b) has shown that if the population statistic is relatively constant, the equal weighting estimator is approximately optimal in terms of a squared error criterion. In order to track time trends more accurately, more weight could be placed on more recent yearly estimates.

Breidt (1999) presents some models that can be used to select the weights for the RA estimator in a more objective manner than a simple arbitrary selection as presented by Roesch and Reams (1999) and Reams and Van Deusen (1999). In order to test the properties of these estimators, we constructed a simulation population with data from two periodic FIA surveys. Survey data were used so that our pseudo-population would have realistic trends and variances for selected variables of interest. We then simulated random samples from this population and compared the simulated estimator distributions for accuracy and precision of the estimators.

Simulation Population

In order to construct a population that possessed attribute values that were realistic, FIA survey data from the cycle 4 and cycle 5 periodic surveys of the Aspen-Birch Unit (ABU) in Minnesota were used as endpoints for the 14 yr span from 1977 through 1990. The ABU is an approximately 8.65 million ac area in northeastern Minnesota of which 86.5% was classified as forest in 1977 (Jakes 1980) and 85% was classified as forest in 1990 (Leatherberry et al. 1995). In both cycle 4 and cycle 5, each plot was sampled using a 10 point variable radius sampling design. Each point was the center for a smaller fixed radius micro plot where trees with a dbh of 1 to 4.9 in. were measured. Trees with a dbh ≥5 were measured on the variable radius plots centered at the points.

To begin construction of the simulation population, individual tree values were used for the plots that were physically measured in both cycles. For all of the simulations, however, plot values of each variable were used. Plot values were calculated by multiplying the tree value by its per-acre expansion factor and summing over all trees on a given plot. These variable radius methods are given in Schreuder et al. (1993, p. 117). The total number of trees on these plots will be denoted by N, and each tree is assigned an identification number $j \in \{1,\ldots,N\}$. Trees were then separated into one of three groups: (1) trees alive at both measurements (survivor trees), (2) trees that were alive at the first measurement but died before the second measurement (mortality trees), and (3) trees that were not present at the first measurement but were alive at the second measurement (in-growth trees). We will denote the subset of survivor trees by N_s. The subsets of mortality and in-growth trees will be denoted by N_m and N_g respectively.

Mortality trees were split into three additional groups based on cause of death. This more detailed classification was designed to better mimic mortality effects on a given plot. The first classification was for trees that could reasonably be modeled as having died randomly and independently at some time during the 13 yr span. The causes of death that were included in this group were suppression, weather damage, and animal damage. The next classification was for trees that should be modeled as having died in the same year for a given cause of death. For example, all trees on a given plot that died of harvest should be modeled as dying in the same year. The causes of death included in this classification were fire, harvest, and land use conversion. It can be argued that weather damage should be included in this classification. However, all weather-damaged trees may not have been damaged by the same weather event on a given plot. Therefore, for ease of construction, trees that died from weather-related causes were placed in the first group. Trees that died from disease or insect damage made up the third mortality classification. These trees were modeled as randomly dying over a 3 yr period on a given plot.

There were very few trees that were not present at the first measurement and were found dead at the second measurement, so, they were excluded from the population for convenience. For this population, the date of "death" for mortality trees is defined to be the last year the tree was observed alive, and the date of "birth" for in-growth trees is defined to be the first year the tree is observed. In the description of the imputation procedure we will use Y_{ij}, $i = 1,\ldots,14$ and $j = 1,\ldots,N$ for either the dbh or volume of tree j in year i; the procedures were the same for both variables. The subscript $i = 1$ will correspond to year 1977, $i = 2$ will correspond to 1978, and so on.

The first step in simulating yearly values for the Y_{ij} was to calculate the growth rate, ξ_j, for each tree. The growth rate for survivor tree $j \in N_s$ was calculated by

$$\xi_j = (Y_{14j} - Y_{1j}) / 13 \cdot$$

Growth rates for mortality and in-growth trees were calculated by averaging the growth rates of survivor trees that were the same species and whose dbh measurement was within 1 inch of the mortality or in-growth tree in 1977 or 1990, respectively.

Next, dates of birth and death were selected for each tree. Survivor trees were defined to have a birth date of $i = 1$ and a death date of $i = 14$. Mortality trees were defined to have a birth date of $i = 1$. For trees in the first mortality classification, the death date was selected uniformly from{1,...,13}. The death date for trees in the plot-wide classification was selected in the same way, but it was applied to all trees that died from a given cause (i.e., fire) on a given plot. A plot death date was randomly selected from {2,...,13} for trees dying from disease or insects. Then, for each tree in the disease class, a random variable selected uniformly on {–1,0,1} was added to the plot date so that disease trees would randomly die over a 3 yr period on any given plot. In-growth trees were defined to have a death date of $i = 14$. The birth date for in-growth trees was randomly selected from {2,...,14} .

Initial and final "alive" values, Y_{bj} and Y_{dj}, were calculated for each tree. The subscripts b and d are used to represent the value of i which is the birth and death date respectively for tree j. The initial and final values for survivor trees were assigned the values recorded during the inventories. The initial value, Y_{bj}, for mortality trees was assigned the cycle 4 inventory value. The final value for tree $j \in N_m$, was calculated in the following fashion:

$$Y_{dj} = Y_{bj} + \xi_j (D_j - 1) \qquad j \in N_m$$

where D_j is the number of years tree j was alive. The final values for in-growth trees were assigned the values recorded in the cycle 5 inventory. The initial value for in-growth tree $j \in N_g$ was calculated by

$$Y_{bj} = Y_{dj} - \xi_j (D_j - 1) \qquad j \in N_g \qquad (1)$$

if it was not located on a microplot. If the calculation in (1) gave an initial measurement <5 in. then the date of birth was adjusted so the initial value was ≥5 in. If tree j was located on a microplot, then the initial dbh was set to 1 in. Initial volume values were then calculated with (1) using the new D_j when it had been adjusted. For microplot trees, initial volume was set to 0.

In order to simulate some noise in yearly Y_{ij} values for tree j, the interval [0,1] was randomly partitioned with $D_j - 2$ random partitions and the end points {0,1}. The values of these points were used to represent the cumulative proportion of total growth, $Y_{dj} - Y_{bj}$, of tree j occurring at "measurement" in year i. Constructing yearly values in this manner forces the property $Y_{i+1,j} \geq Y_{ij}$ for all trees and all years. In order to carry out this method of construction for tree j, $D_j - 2$ ordered random variables, $U_{b+1,j},\ldots,U_{d-1,j}$, with distribution Unif(0,1) were selected. The endpoints, $U_{bj} = 0$ and $U_{dj} = 1$, are added to the collection of random variables. Then, yearly values were calculated by

$$Y_{ij} = U_{ij}(Y_{dj} - Y_{bj}) + Y_{bj} \qquad i = b,\ldots,d$$

for all j. For values of i outside $\{b,\ldots,d\}$, Y_{ij} was set to 0. Also, the volume value for tree j in year i was set to 0 if the corresponding dbh was <5 in. For a given j, the expected measurement for tree j is

$$E[Y_{ij}] = (Y_{dj} - Y_{bj})\left(\frac{i-b}{d-b}\right) + Y_{bj} \qquad i = b,\ldots,d$$

which is the linear interpolation between Y_{bj} and Y_{dj} at year i. In addition,

$$Corr[Y_{ij}, Y_{kj}] = \frac{(i-b)(d-k)}{\sqrt{(i-b)(k-b)(d-i)(d-k)}}, \qquad i < k \in \{b+1,\ldots,d-1\}$$

for a given tree j. Correlations involving Y_{bj} or Y_{dj} are defined to be 0 because they are constants. This correlation function is decreasing as |k – 1| becomes larger.

In the cycle 5 inventory, plots from the cycle 4 inventory that were classified as disturbed were re-sampled with probability 1, while undisturbed plots were re-sampled with probability 1/3 (Leatherberry et al. 1995). In order to correct for this unequal weighting, all undisturbed plots were duplicated twice and added back into the constructed population. So, for every undisturbed plot in the original constructed population, two more were added that possessed the same tree values. This duplicated population is a better representation of the distribution of tree values in the real population. The resulting number of plots in the simulation population was 5,911. Volume trends in the simulation population closely mimicked the published attributes for the ABU (Leatherberry et al. 1995).

Breidt's Estimators

The sampling plan that all of Breidt's (1999) estimators use is a two-phase design. First a sample s of n plots is drawn from area $\mathscr{A}$ of size A according to a probabilistic design $p(\bullet)$. Here we will present results for simple random sampling (SRS) and stratified simple random sampling (STS). The sample, s, is then randomly partitioned in to T subsamples s_{t-T+h}, $h = 1,...,T$, where the partitioning is independent of s. One of these subsamples is carried out in each of the T years, so, the entire primary sample s is measured after T years. In this article, the assumption is also made that the subsamples are of equal size and that all plots have the same probability of being placed in any given subsample, $1/T$. This assumption is not critical, however, and will be discussed shortly. In the following descriptions of these estimators we will consider all population parameters, θ_t, as averages per unit area (e.g. average volume per acre). The extension to population totals is easily accomplished by multiplying the estimator by A.

To begin, using two-phase sampling methodology, θ_{t-T+h}, $h = 1,...,T$, can be unbiasedly estimated by

$$\hat{\theta}_{t-T+h} = A^{-1} \sum_{k \in s_{t-T+h}} \sum_{i=1}^{n_k} \frac{z_{ik}(t-T+h)}{\pi_{ik}(1/T)} \qquad (2)$$

where n_k is the number of trees sampled on plot k, z_{ik} is the variable of interest from the ith tree sampled on plot k, and π_{ik} is the Horwitz-Thompson expansion factor. Under an SRS design, $\pi_{ik} = na_{ik}/A$, where a_{ik} is the tree to unit area expansion factor as described by Eriksson (2001a). For an STS design, $\pi_{ik} = n_j a_{ik}/A_j$, $k \in \mathscr{A}_j$, $j = 1,...,J$, where n_j is the total number of plots sampled in stratum $\mathscr{A}_j$, and A_j is the size of $\mathscr{A}_j$. The problem with these estimators is that with only n/T plots sampled in any given year, they may have an unacceptable level of precision. In future descriptions, we will refer to this unbiased estimator as the SA estimator, $\hat{\theta}_{SA,t-T+h}$.

In practice, equal subsample size is unlikely to be obtained. However, if partition of s into $\{s_{t-T+h}: h = 1,...,T\}$ is completed independently of s and the probability that plot k is included in s_{t-T+h} equals P_{t-T+h}, $h = 1,...,T$, then, $(1/T)$ can be replaced by P_{t-T+h} in $\hat{\theta}_{t-T+h}$, and the estimator will still be unbiased. A small amount of variation in the subsampling proportions would probably not greatly affect the results obtained in this study. A large variation may, however, have an effect on the results depending on the weights assigned to the different subsamples. These effects will not be explored here.

In order to raise the level of precision it has been suggested that information from past years be used in the estimation of parameters for the present year. Therefore, the estimators considered here all have the form

$$\hat{\theta}_{\cdot,t} = \sum_{h=1}^{T} \sum_{k \in s_{t-T+h}} \sum_{i=1}^{n_k} w_{ik} z_{ik}(t-T+h) \qquad (3)$$

where w_{ik} is a weight that reflects the design properties of the sample or an assumed model about the population. In addition, with a little algebraic manipulation, (3) can be rewritten to be the weighted sum of the unbiased estimators, (2). It is easily seen that if identical weights are chosen for all variables, then estimates are internally consistent (the estimate of a sum of variables equals the sum of the separate estimates for each variable).

In describing the estimators, we will divide them into three classes. The first class of estimators is based on the time series ARIMA(0,1,1) model (Brockwell and Davis 1987, p. 274). This class of models includes the equal weighted simple moving average (a simple average of unbiased yearly estimates). The second class is based on the ARIMA(0,2,2) time series model (Brockwell and Davis 1987, p. 274). Finally, the third class is based on a locally weighted least-squares regression prediction (Neter et al. 1989, p. 400).

ARIMA(0,1,1)-Based Estimators

Estimators of θ_t in this class are based on the ARIMA(0,1,1) model

$$\begin{aligned} \hat{\theta}_t &= \theta_t + \epsilon_t, \qquad \{\epsilon_t\}\; iid \;\; N(0,\sigma^2\} \\ \theta_t &= \theta_{t-1} + \eta_t, \quad \{\eta_t\}\; iid \;\; N(0,\sigma_\eta^2\} \end{aligned}$$

For example, the average total volume over the state of Minnesota next year is equal to this year's average total volume plus some random noise that is independent from year to year. In addition, the ARIMA(0,1,1) model assumes that you have an unbiased estimate of average total volume in each year.

The best mean square estimator of θ_t derived from this model is

$$\hat{\theta}_{TS1,t} = \sum_{h=1}^{T} w_{t-T+h} \hat{\theta}_{t-T+h}$$

where the weights, w_{t-T+h}, $h = 1,...,T$, are given by the last row of the matrix

$$I - \Delta_1' \left[\frac{\sigma_\eta^2}{\sigma^2} I + \Delta_1 \Delta_1' \right]^{-1} \Delta_1$$

The matrix Δ_1 is the $T \times (T-1)$ first difference matrix given by

$$\Delta_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

and I is an appropriately sized identity matrix. The parameter σ_n^2/σ^2 is usually called the "signal to noise ratio" (SNR). As the SNR increases from 0 to ∞, an increasing amount of weight is placed on more recent conditions. These estimators will be denoted by $TS1(SNR = x)$ for an ARIMA(0,1,1) based estimator with an $SNR = x$.

ARIMA(0,2,2)–Based Estimators

This estimator class is based on the ARIMA(0,2,2) model

$$\hat{\theta}_t = \theta_t + \in_t, \qquad \{\in_t\} \; iid \;\; N(0,\sigma^2)$$

$$\theta_t = \theta_{t-1} + \beta_t + \eta_t, \quad \{\eta_t\} \; iid \;\; N(0,\sigma_\eta^2)$$

$$\beta_t = \beta_{t-1} + \xi_t, \qquad \{\xi_t\} \; iid \;\; N(0,\sigma_\xi^2)$$

For example, next year's average total volume in Minnesota is equal to this year's average total volume plus some noise that is dependent between years and some more noise that is independent between years. This model also assumes that there is an unbiased estimator of average total volume each year. The resulting estimator, $\hat{\theta}_{TS2,t}$, has the same form as $\hat{\theta}_{TS1,t}$. However, the weights are generated by the last row of the matrix

$$I = \Delta_2' \left[\frac{\sigma_\xi^2}{\sigma^2} I + \frac{\sigma_\eta}{\sigma^2} \Delta_1 \Delta_1' + \Delta_2 \Delta_2' \right]^{-1} \Delta_2$$

The matrix Δ_2 is the $T \times (T-2)$ second difference matrix given by

$$\Delta_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}$$

Also, Δ_1 is the same as (4) except the last row has been removed. Here we will consider the *SNR* to be the vector $(\sigma_\xi^2 / \sigma^2, \sigma_\eta^2 / \sigma^2)$. Once again, as the *SNR* becomes larger, more weight is placed on current conditions. This is easy to see, as either element of the *SNR* tends to ∞, the variation in the population parameter becomes much greater than the variance of the estimator. Therefore, the population is changing rapidly and past data has little information about current conditions. The ARIMA(0,2,2) based estimator will be denoted by $TS2(SNR = x,y)$.

Locally Weighted Least-Squares Estimators

The final class of estimators is based on a plot-level model. Let $y_k(t - T + h)$ be the average per unit area of the variable $z(t - T + h)$ in the plot centered at point $k \in \mathscr{A}$ in year $t - T + h$. For both *SRS* and *STS* designs

$$y_k(t - T + h) = \sum_{i=1}^{n_k} z_{ik}(t - T + h) / a_{ik}, k \in s_{t-T+h}.$$

The plot level expansion factors, π_k, are given by A/n for *SRS* designs and A_j/n_j for STS designs. The plot level model considered is

$$y_k(t - T + h) = m(t - T + h) + \delta_k + \in_k (t - T + h), \qquad h = 1,\ldots,T)$$

where $m(t - T + h)$ is a smooth deterministic function, $\{\delta_k\}$ is a set of *iid* r.v.'s with mean 0 and variance σ_δ^2 and $\{\in_k(t - T + h)\}$ is a zero mean stochastic process. So, $\theta_{t-T+h} = E[y_k(t - T + h)] = m(t-T+h)$. Over the T year window of plot rotation, the function $m(\bullet)$ can be approximated by a polynomial of order q. Therefore, a natural estimator of $m(\bullet)$ is the predicted value of a linear regression of $y_k(t)$ versus year. If we define the $n \times 1$ vector of all plots measured arranged according to the year in which they were sampled

$$\mathbf{y} = \left[\left[y_k(t - T + h) \right]_{k \in s_{t-T+h}} \right]_{h=1}^T$$

the $n \times q + 1$ matrix of years in which each plot was measured

$$\mathbf{X} = \left[\left[1 \;\; h \;\; h^2 \cdots h^q) \right]_{k \in s_{t-T+h}} \right]_{h=1}^T$$

and the $n \times n$ matrix of plot selection probabilities

$$\mathbf{W} = \left[\text{block diag} \left\{ \text{diag} \left\{ \frac{1}{\pi_k} \right\}_{k \in s_{t-T+h}} \right\} \right]_{h=1}^T$$

then, the locally weighted linear regression estimator of θ_t is given by

$$\hat{\theta}_{LL,t} = \left[1 \;\; T \cdots T^q \right] \left[\mathbf{X'WX} \right]^{-1} \mathbf{X'Wy}$$

In both SRS and STS designs, $\hat{\theta}_{LL,t}$ can be rewritten to give the following form

$$\begin{aligned} \hat{\theta}_{LL,t} &= \left[1 \;\; T \cdots T^q \right] \left[\mathbf{X'WX} \right]^{-1} \mathbf{X'Wy} \\ &= \sum_{h=1}^{T} \left(\frac{A}{T} \right) \left[1 T \cdots T^q \right] \left[\mathbf{X'WX} \right]^{-1} \begin{bmatrix} 1 \\ h \\ \vdots \\ h^q \end{bmatrix} \\ &\quad \times \left[A^{-1} \sum_{k \in s_{t-T+h}} \sum_{i=1}^{n_k} \frac{z_{ik}}{\pi_{ik}(1/T)} \right] \\ &= \sum_{h=1}^{T} w_{t-T+h} \hat{\theta}_{t-T+h} \end{aligned}$$

We will denote this estimator by *LL*.

Special Cases, Connections, and Considerations

The first consideration applies to use of both *TS*1 and *TS*2 estimators. Notice, the weights reflect properties of the variable of interest through the *SNR*, which will be unknown in any forest survey. So, in practice, the *SNR* for both *TS*1 and *TS*2 estimators is usually set to a fixed constant based on the level of smoothing desired. In addition, there is a connection between the *TS*1 estimators and the simple moving average, which we will denote by *MA*. The $TS1(SNR = 0)$ estimator is equivalent to the equal weight simple moving average.

The second consideration involves the *LL* estimator. Notice, the weights reflect only the sample design. The weights are the same for any variable measured within a given sampling design, so, there is no decision to be made about smoothing parameters. Also, if both elements of the *SNR* in the *TS*2 estimators are set to 0, then, for the *SRS* design $\hat{\theta}_{LL,t} = \hat{\theta}_{TS2,t}$. For this reason, we will group the *LL* estimator with the *TS*2 class. Also, if the order of the regression model is set to $q = 0$, the *LL* estimator is the same as the *MA* estimator.

None of the estimators derived here should be considered to have better performance *a priori*. The simulation population was constructed by a very nonparametric method and is not a special case of any of the models presented for estimator derivation. We felt that constructing the population without a parametric model better mimics a real forest population in the fact that truth never fits a probabilistic model perfectly.

Efficiency and Sample Size Illustration

A simple population model can provide a brief illustration of the approximate, relative efficiency of the different rolling sample estimators of the current value of the population parameter. To begin, suppose that plot measurements follow the model:

$$Y_{it} = \alpha + \beta(t) + \epsilon_{it} \tag{5}$$

where Y_{it} is the measurement for plot *i* at time *t*, and the errors (ϵ_{it}) are independent with respect to plots, have mean 0 and constant variance over time, σ^2. If a first order model for $\hat{\theta}_{LL,t}$ is used, the population size is large, plots are sampled with equal probability, and plot subsample size equals n/T, then, the following mean square error approximations can be easily derived for the yearly unbiased (*SA*), moving average (*MA*), and local linear (*LL*) estimators:

$$MSE\left[\hat{\theta}_{SA,t}\right] \approx T\sigma^2 / n$$

$$MSE\left[\hat{\theta}_{MA,t}\right] \approx (T-1)^2\beta^2 / 4 + \sigma^2 / n$$

and

$$MSE\left[\hat{\theta}_{LL,t}\right] \approx (4T+2)\sigma^2 / \{n(T+1)\}$$

In addition, for both the *TS*1 and *TS*2 estimators with fixed SNR or any general rolling average estimator

$$MSE\left[\hat{\theta}_{RA,t}\right] \approx \left(\sum_{h=1}^{T} h w_{t-T+h} - T\right)^2 \beta^2 + T\sigma^2 n^{-1}\left(\sum_{h=1}^{T} w_{t-T+h}^2\right)$$

where, w_{t-T+h}, $h = 1,...,T$, are the weights given to the yearly unbiased estimates.

If the relative efficiency of estimators is measured by the ratio of *MSE*s then, the relative efficiency of the *MA* estimator compared to the *LL* estimator is approximately

$$n\frac{(T-1)^2(T+1)}{8(2T-1)}\frac{\beta^2}{\sigma^2} + \frac{T+1}{2(2T-1)}. \tag{6}$$

So, if there is no trend to the data, then the *LL* estimator can never be as efficient as the *MA* estimator. If there is even the slightest trend, however, it is theoretically possible (although maybe not practically possible) to take a large enough sample such that the *LL* estimator is more efficient. This is due to the fact that the bias of the *MA* estimator is not a function of sample size, while for this simple population the *LL* estimator is unbiased. Using (6), the sample size, n^*, required to attain equal efficiency between the *MA* and *LL* estimators is

$$n^* \approx \frac{12}{(T-1)(T+1)}\frac{\sigma^2}{\beta^2}. \tag{7}$$

One can see that n^* essentially represents a trade-off between the trend of a population and the variance of plot measurements.

All of the derivations provided in this section were based on the simple population model (5). The extent to which these results will generalize to more complicated populations is unknown. The results shown are intended to provide a heuristic feel for how trend and variance combine to influence sample size and efficiency considerations for the rolling average estimators.

Simulations

For this simulation, we will use two types of variables that are often of interest to analysts. The first type is measurements of wood volume. In this study, we simulated use of the estimators on: average volume per acre, average mortality volume per acre, and average on-growth volume per acre. In addition, average volume per acre for nine species was also simulated. For the experiment, we chose three common species, three moderately common species, and three rare species. The common species included: quaking aspen (*Populus tremuloides*), paper birch (*Betula papyrifera*), and balsam fir *(Abies balsamea*). The moderately common species were red pine (*Pinus resinosa*), white spruce (*Picea glauca*), and sugar maple (*Acer saccharum*). The rare species were cottonwood (*Populus deltoides*), black willow (*salix nigra*), and silver maple (*Acer saccharinum*). The next type of variable is condition class. For this type, we simulated use of the estimators on proportion of area in sawtimber, poletimber, and sapling status. These classifications were generated using a simplification of the algorithm used by the North Central Region FIA. Instead of applying the algorithm to each sample point of a plot, it was applied to the whole plot.

Simulations with the pseudopopulation were accomplished by first selecting 1,000 plots by simple random sampling with replacement from the simulated plots. The reason for sampling with replacement was to simulate sampling from a region with area *A* whose response surface for the variables listed previously is a step function. Each of the steps has the height of one of the simulated plots and an area of *A*/5911 ac. Sampling without replacement is equivalent to randomly locating a sample point somewhere in this region. In addition, using the formulas in Ripley (1981), expectations and variances for the variables mentioned previously are calculated simply by using the means and variances of the 5,911 simulated plots. After the initial sample was taken, each plot in the sample was randomly assigned to 1 of 5 subsamples of size 200. The same 1,000 plots were used throughout the 14 yr period 1977,...,1990. So, once year 6 was reached, the plots from the first subsample were used again. Following selection of the samples, the subsamples were used to calculate unbiased estimates of population parameters with the two phase unbiased estimator [*SA*, (2)] for each year, 1977,...,1990. Then, several of Breidt's weighted estimators were used to estimate the current status for the years 1981,...,1990. For on-growth and mortality volume, the year indices were 1978,...,1990 and 1982,...,1990. In each simulation, this procedure was replicated 5,000 times. We believed that this number of replications was adequate based on the fact that the same estimators in different simulations behaved very similarly.

In order to use the time-series-based estimators, we made some practical decisions concerning the *SNR*s. The first decision that was made was to set the first *SNR* in the *TS*2(*SNR* = *x*,*y*) estimator to be 0. We thought that modeling the population as having a constant mean rate of increase or decrease over a 5 yr period was a sufficient approximation and left only one *SNR* to be chosen. In order to calculate optimal *SNR*s, we needed to calculate σ_η^2 for both models. In each model, $\mathrm{Var}[\theta_t - \theta_{t-1}] = \mathrm{Var}[const. + \eta_t]$. Therefore, using the notation from the simulation population description, we calculated σ_η^2 by

$$\sigma_\eta^2 = Var\left\{\overline{Y}_i - \overline{Y}_{i-1} \;:\; i \in \;1978,...,1990\right\},$$

where $\hat{Y}_t$ is the population mean of variable *Y* in year *t*. The second parameter in the *SNR*s, σ^2, is simply the sampling variance of the two-phase estimator and was calculated according to Eriksson (2001a). All of the *SNR*s were less than 0.2, and many were very close to 0. Thus, the *TS*1 and *TS*2 estimators with optimal *SNR*s will be very close to the *MA* and *LL* estimators, respectively. We suspect that the true *SNR*s will be less than 1 for almost any forest resource, since change over time is relatively small compared to differences over, say, a county. To mimic the selection of constant *SNR*s, we chose four values for the *SNR*s in the interval 0 to 1. We will denote a time series based estimator with an optimal *SNR* by *TS*1(*SNR* = *opt*) or *TS*2(*SNR* = 0,*opt*).

There were five simulations that were conducted. The first simulation involved the volume variables: average volume per acre, average on-growth volume per acre, and average mortality volume per acre. The estimators used were: *MA*, *LL*, *TS1(SNR* = *opt*), *TS*1(*SNR* = 1), *TS*2(*SNR* = 0,1), and *TS*2(*SNR* = 0,*opt*). The second simulation used the same volume variables but the estimators tested were *MA*, *LL*, *TS1*(*SNR* = 0.25), *TS*1(*SNR* = 0.5), *TS*1(*SNR* = 1), *TS*2(*SNR* = 0,0.25), *TS*2(*SNR* = 0,0.5), and *TS*2(*SNR* = 0,1). This simulation was performed to see how sensitive these estimators are to the choice of *SNR*s. The third simulation involved the species domains of average volume per acre. The estimators used were *MA*, *LL*, *TS*1(*SNR* = 1), and *TS*2(*SNR* = 0,1). The purpose of this simulation was to determine how these estimators behave when they are used to estimate domains of a general variable. The fourth and fifth simulations were the same as the first and second simulations except the classification variables were used.

We used two criteria to judge the quality of a particular estimator: root mean squared error (*RMSE*) and the bias ratio (*BR*). For the estimator σ_η^2 of the population value θ_t, the simulated *RMSE* and *BR* were calculated by

$$RMSE(\hat{\theta}_t) = \sqrt{\frac{1}{r}\sum_{i=1}^{r}(\hat{\theta}_{t,i} - \theta_t)^2}$$

and

$$BR(\hat{\theta}_t) = \frac{\left|\left(\sum_{i=1}^{r} \hat{\theta}_{t,i}\right) / r - \theta_t\right|}{S(\hat{\theta}_t)}$$

where *r* is the number of simulated samples drawn, $\hat{\theta}_{t,i}$ is the estimate of θ_t for sample *i*, and $S(\hat{\theta}_t)$ is the sample standard deviation formula applied to the *r* values of $\hat{\theta}_{t,i}$. The *BR* statistic was chosen to determine if bias might play a large role in the actual coverage of a nominal confidence interval (CI). A value of $|BR| < 0.1$ implies the bias effect on the CI can essentially be ignored and even for a value <0.5, the bias effect is not pronounced (Sarndal et al. 1992).

Results

The main result from these simulations was that all of the *RA* estimators outperformed the *SA* estimator in every simulation, with every variable with regards to *RMSE*. This result makes intuitive sense. There was, however, a definite gradient among the estimators in terms of *RMSE* and *BR*. In many cases, low average *RMSE* was accompanied by a high average *BR*. In terms of average *RMSE*, the *TS*1 class of estimators outperformed the *TS*2 estimators and the *LL* estimator in every simulation. In addition, the *MA* estimator performed the best in most of the simulations. The largest increase in average *RMSE* from the best estimator to the *MA* estimator was only 7.31%.

In the first simulation, where average total, on-growth, and mortality volume were estimated, *TS*1(*SNR* = *opt*) had the lowest average *RMSE* for volume and on-growth, while the *MA* estimator had the lowest average *RMSE* for mortality volume (Table 1). The *TS*1(*SNR* = *opt*) and *MA* estimators performed almost identically in estimating the three parameters, the largest percent difference in average *RMSE* being 2.84%. Figure 1 displays the performance of the first simulation estimators in terms of bias and variance. The *MA* estimator had the tightest one standard deviation (SD) band, and the true value for each variable was usually within the band.

In the second simulation, where average total, on-growth, and mortality volume were estimated again, the *MA* estimator always had the lowest average *RMSE* (Table 2). Figure 2

Table 1. Results of the first simulation. Rows are arranged in ascending order according to average RMSE within each variable.

Variable	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Volume	TS1(SNR = *opt*)	25.39	0.314	0
	MA	25.41	0.316	0
	TS1(SNR = 1)	35.85	0.077	0
	LL	41.28	0.035	0
	TS2(SNR = 0,*opt*)	41.28	0.035	0
	TS2(SNR = 0,1)	43.15	0.023	0
	SA	53.20	0.010	0
On-growth	TS1(SNR = *opt*)	1.71	0.539	4
	MA	1.76	0.803	5
	TS1(SNR = 1)	2.05	0.212	1
	LL	2.33	0.171	1
	TS2(SNR = 0,*opt*)	2.34	0.153	1
	TS2(SNR = 0,1)	2.42	0.102	0
	SA	2.96	0.012	0
Mortality	MA	3.65	0.386	3
	TS1(SNR = *opt*)	3.78	0.272	1
	TS1(SNR = 1)	4.95	0.125	1
	TS2(SNR = 0,*opt*)	5.96	0.252	2
	LL	6.02	0.277	2
	TS2(SNR = 0,1)	6.02	0.164	0
	SA	7.26	0.008	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

illustrates the performance of the *TS*1 estimators through time. As the *SNR* becomes larger, more weight is placed on current conditions and the 1 *SD* band becomes wider. In Table 2 it can be seen that the *TS*2 class of estimators seems to be fairly insensitive to changes in the *SNR*. There was only at most a 4.56% difference among average *RMSE* values for the *TS*2 class estimators.

Once again, in the third simulation, where average volume for three species domains was estimated, the *MA* estimator had the lowest average *RMSE* over all species (Tables 3–5). There was also almost no detectable bias problem for the *MA* estimator in each species. In all but one year, $BR < 0.5$ for all species. Figures 3 and 4 show that the *MA* estimator performed well for the common and moderately common species. It had the tightest 1 *SD* band, and the bias was small. In the case of the rare species domain, Figure 5 illustrates that the *MA* variance can be higher than the other estimators if the variable decreases to 0. It also illustrates a potential problem with the *TS*2 class of estimators, namely negative estimates. Since the *TS*2 estimators can have negative weights, negative estimates are possible. As Cottonwood and Black willow increase from 0, the *MA* estimator once again performed very well.

In the fourth simulation, where proportion of area classified by one of three different stand size classes was estimated, *TS*1(*SNR* = *opt*) performed the best in terms of average *RMSE* (Table 6). The greatest difference, however, between the *MA* estimator and the *TS*1(*SNR* = *opt*) was 4.21%. Within a given year, the small variance of the classification variables created a large average bias ratio. The *BR* was > 0.5 in almost every year for sawtimber and poletimber. For the first two years, *BR* > 2 for the sapling class.

In the fifth simulation, the *MA* estimator performed nearly the best for each size class; however, bias problems were experienced due to low population variability when com-

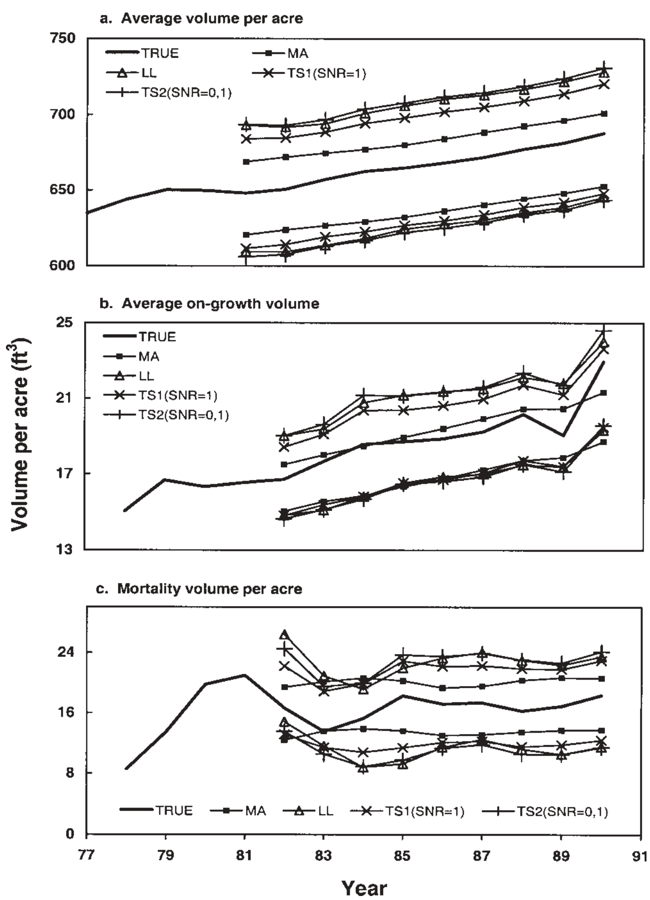

Figure 1. Estimator performance in the first simulation (volume variables). The lines represent a 1 SD band for a given estimator over 5,000 simulated samples. The *TS*1(*SNR* = *opt*) and *TS*2(*SNR* = 0,*opt*) were excluded because they were virtually identical to the *MA* and *LL* estimators.

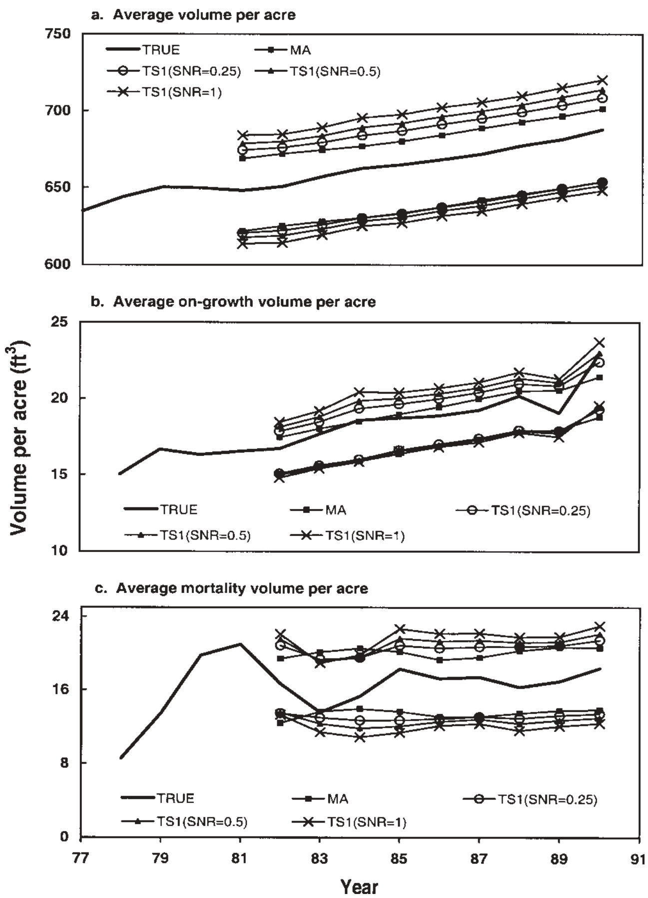

Figure 2. Performance of the *TS*1 estimators in the second simulation (volume variables). The lines represent a 1 standard deviation band for a given estimator over 5,000 simulated samples.

pared to the population trend. Figure 6 shows that the 1 *SD* band of the *MA* estimator drifts away from the true population trend. When exploring different fixed *SNR* values, the *TS*1(*SNR* = 0.25) performed the best with regards to average *RMSE* (Table 7). The *MA* estimator had the second lowest *RMSE* with the exception of the poletimber classification, where *TS*1(*SNR* = 0.5) had the second lowest average *RMSE*. The *MA* estimator had the smallest 1 SD band, but, it tended to stray from the true population value. Figure 7 shows that for each variable, the true value was almost always contained within the 1 *SD* band of the *TS*1(*SNR* = 0.25) estimator. By applying slightly more weight to current estimates, the bias problems of the *MA* [*TS*1(*SNR* = 0) seem to disappear. In addition, Table 7 shows that once again, the performance of the *TS*2 class of estimators seems to be rather robust to changes in the value of the *SNR*. The largest difference in average *RMSE* for the *TS*2 estimators was 5.07%.

Discussion

Certainly, we have shown for this simulation population that the *TS*1 class of estimators, which contains the *MA* estimator, performed the best in terms of *RMSE* for our simulation population. This class can, however, have bias problems. Although, if the *SNR* was high enough for the *TS*1 estimator, bias problems were often be eliminated at the expense of *RMSE*. Another conclusion that can be drawn from these simulations is that the *MA* estimator performs well. At first, simply averaging yearly estimates does not seem like it would produce reliable estimates of current status. If spatial variation is high enough, however, the *MA* estimator seems to work well, even under some moderate

Table 2. Results of the second simulation (*TS* estimators). Rows are arranged in ascending order according to average RMSE within each variable.

Variable	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Volume	MA	24.60	0.292	0
	TS1(SNR = 0.25)	27.46	0.156	0
	TS1(SNR = 0.5)	30.94	0.100	0
	TS1(SNR = 1)	35.51	0.060	0
	LL	41.03	0.035	0
	TS2(SNR = 0,0.25)	41.30	0.031	0
	TS2(SNR = 0,0.5)	41.81	0.027	0
	TS2(SNR = 0,1)	42.99	0.023	0
	SA	53.28	0.016	0
On-growth	MA	1.71	0.782	5
	TS1(SNR = 0.25)	1.71	0.453	2
	TS1(SNR = 0.5)	1.84	0.310	1
	TS1(SNR = 1)	2.04	0.192	1
	LL	2.33	0.178	1
	TS2(SNR = 0,0.25)	2.34	0.154	0
	TS2(SNR = 0,0.5)	2.36	0.135	0
	TS2(SNR = 0,1)	2.42	0.108	0
	SA	2.98	0.014	0
Mortality	MA	3.64	0.389	2
	TS1(SNR = 0.25)	3.92	0.245	1
	TS1(SNR = 0.5)	4.34	0.185	0
	TS1(SNR = 1)	4.91	0.126	0
	TS2(SNR = 0,0.25)	5.90	0.235	1
	TS2(SNR = 0,0.5)	5.90	0.205	1
	LL	5.96	0.273	2
	TS2(SNR = 0,1)	5.98	0.162	0
	SA	7.21	0.003	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

Table 3. Results of the third simulation (common species). Rows are arranged in ascending order according to average RMSE within each species.

Species	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Quaking aspen	MA	11.58	0.066	0
	TS1(SNR = 1)	17.50	0.025	0
	LL	20.32	0.031	0
	TS2(SNR = 0,1)	21.24	0.021	0
	SA	26.24	0.010	0
Paper birch	MA	6.84	0.068	0
	TS1(SNR = 1)	10.35	0.021	0
	LL	12.00	0.028	0
	TS2(SNR = 0,1)	12.57	0.019	0
	SA	15.56	0.010	0
Balsam fir	MA	5.22	0.192	0
	TS1(SNR = 1)	7.60	0.050	0
	LL	8.80	0.028	0
	TS2(SNR = 0,1)	9.19	0.018	0
	SA	11.35	0.013	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

Table 4. Results of the third simulation (moderately common species). Rows are arranged in ascending order according to average RMSE within each species.

Species	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Sugar maple	MA	3.84	0.092	0
	TS1(SNR = 1)	5.74	0.017	0
	LL	6.23	0.013	0
	TS2(SNR = 0,1)	6.94	0.010	0
	SA	8.56	0.007	0
Red pine	MA	4.74	0.171	0
	TS1(SNR = 1)	7.02	0.040	0
	LL	8.07	0.020	0
	TS2(SNR = 0,1)	8.46	0.018	0
	SA	10.44	0.014	0
White spruce	MA	2.32	0.251	0
	TS1(SNR = 1)	3.42	0.053	0
	LL	3.93	0.021	0
	TS2(SNR = 0,1)	4.13	0.016	0
	SA	5.12	0.011	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

Table 5. Results of the third simulation (rare species). Rows are arranged in ascending order according to average RMSE within each species.

Species	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Cottonwood	MA	0.035	0.345	0
	TS1(SNR = 1)	0.049	0.124	0
	LL	0.059	0.102	0
	TS2(SNR = 0,1)	0.060	0.082	0
	SA	0.070	0.009	0
Black willow	MA	0.036	0.112	1
	TS1(SNR = 1)	0.058	0.058	0
	LL	0.067	0.051	0
	TS2(SNR = 0,1)	0.071	0.041	0
	SA	0.086	0.011	0
Silver maple	MA	0.57	0.057	1
	TS1(SNR = 1)	0.83	0.017	0
	LL	0.97	0.014	0
	TS2(SNR = 0,1)	1.00	0.013	0
	SA	1.23	0.013	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

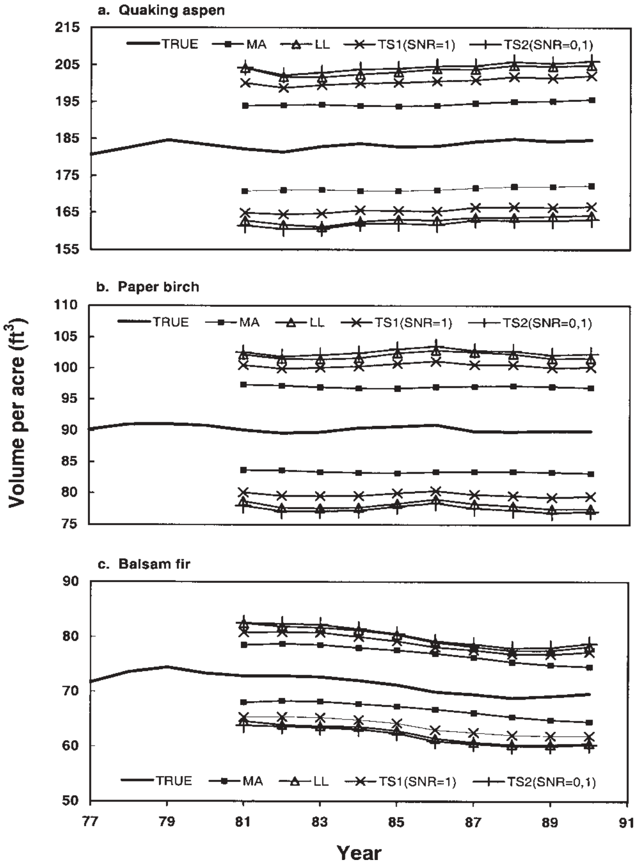

Figure 3. Estimator performance on the common species domains in the third simulation. The lines represent a 1 standard deviation band for a given estimator over 5,000 simulated samples.

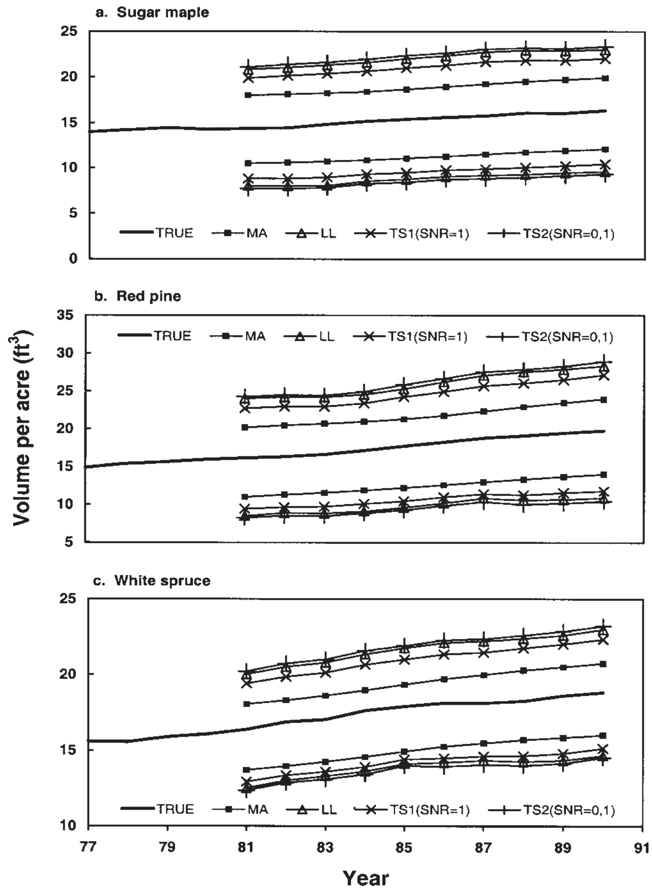

Figure 4. Estimator performance on the moderately common species domains in the third simulation. The lines represent a 1 standard deviation band for a given estimator over 5,000 simulated samples.

temporal trends. Even if it is desirable to choose a less biased *TS*1 estimator, it would be wise to choose a constant *SNR*, at least within each variable type. This would make estimation much less complicated as well as preserving additivity among variable domains.

We suspect that any tree measurement variables are likely to be similar in variation structure to the volume related variables we simulated. Therefore, the *MA* estimator would be a good choice for these type of variables. It has a low variance and would provide the stability of estimates from year to year. The *MA* estimator will react slowly to sudden changes in tree attribute variables. We feel, however, that sudden trend changes in these attributes are unlikely without an obvious catastrophic event such as a large-scale weather event or disease epidemic, in which case it would probably not be advisable to continue sampling and estimation procedures in the usual fashion. In addition, by using an estimator that tracks the current conditions more closely, you may create apparent small trends due to its higher variance. Another consideration is the length of the cycle (T). We simulated sampling with a 5 yr cycle. Longer cycles will certainly have negative effects on the performance of the *MA* estimator in terms of bias.

The smaller within-year variance of condition class variables tends to exaggerate bias effects. Therefore, even though the *MA* estimator had nearly the lowest average *RMSE*, it may be advisable to choose an $SNR > 0$ to reduce these bias effects. How to choose the *SNR* is unclear. It may seem that arbitrary selection of the *SNR* is no better than arbitrary selection of the weights themselves; however, the *SNR* represents an easily interpretable parameter in terms of variation over time versus spatial variation, whereas the weights themselves do not have an interpretation involving the actual population. For example, all of the simulated variables in this study had *SNR*s less than 0.2. Therefore, variation in average yearly change for each variable was less than 20% as large as the corresponding average spatial variation for each variable.

Another consideration is that these simulations were conducted in a very simple fashion without the use of auxiliary information. In recent years, it has been suggested that surveys should make use of auxiliary information to reduce estimator variance or obtain a higher sample in each year (Czaplewski 1999, Robinson et al.1999). The first consideration is how to incorporate the auxiliary information into the *RA* estimators. In the case of the *LL* estimator, incorporation of the information is easy. All that needs to be done is to place the information for each tree sampled into the **X** matrix just as in ordinary regression analysis. By adding the extra information, the *LL* estimator may compete better with the *TS*1 estimators in terms of *RMSE*. Another consideration is how the *MA* estimator will fare once auxiliary information is used. The

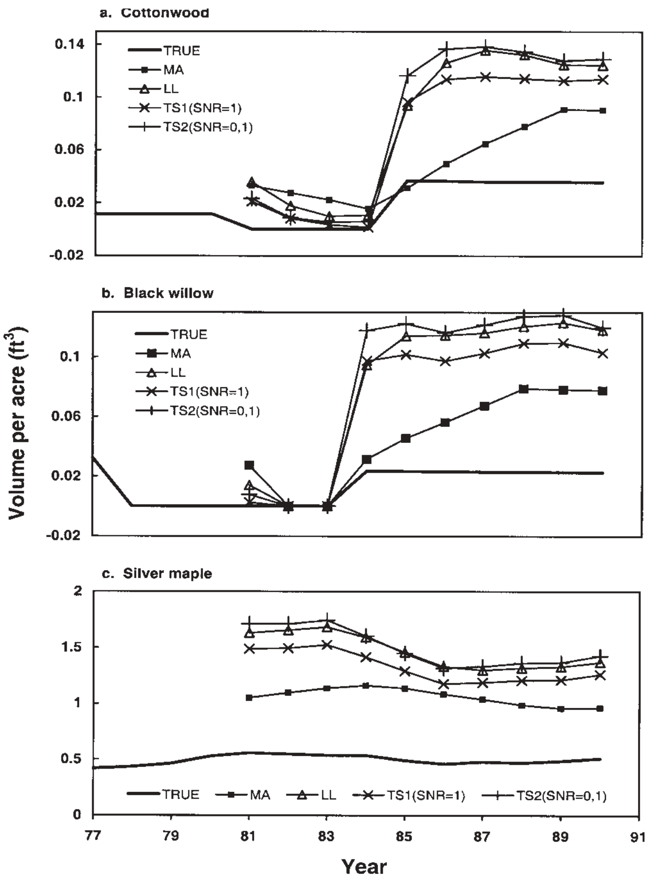

Figure 5. Estimator performance on the rare species domains in the third simulation. The lines represent a 1 standard deviation upper bound for a given estimator over 5,000 simulated samples. The lower bound for *TS*1 estimators is 0 and the lower bound for *TS*2 and *LL* estimators is <0.

reason that the *MA* estimator performs so well is that the variance of the *SA* estimator within a given year is usually very large. If that variance could be substantially reduced with auxiliary information, the *MA* estimator may not perform as well compared to other *TS*1 estimators.

The *TS*2 estimators, including the *LL* estimator, do not perform as well in terms of *RMSE* because they fit the data too closely. Thus, for populations with low *SNR*s, they tend to track noise rather than any real trend. While this produces weighted estimators that are nearly unbiased, the variance of these estimators is high. The high variance often produced average *RMSE* that were not much better than the *SA* estimator, which we considered the upper limit. So, even though these estimators had good bias properties, their large variability made them less desirable for general use in FIA surveys.

Overall, drawing conclusions about optimal or best estimators in a multiresource survey is usually not possible, and when the survey is drawn out over time, it is even more difficult. For the population studied here, the amount of trend in relation to the variance of the individual panel estimators (i.e., the signal to noise ratio) was just too high for the more complex estimators to perform well. If the sample size were increased for a variable with low spatial variation, the *TS*2 estimators, including *LL*, might perform more competitively. For instance, applying Equation (7) to the sawtimber class variable, $n^* \approx 3000$. So, the sample size in the simulation would have to be tripled for the performance of the *LL* estimator to equal the *MA* estimator. However, for a variable like total volume per acre, $n^* \approx 35{,}000$. Thus, increasing sample size to the maximum that is practically feasible is not likely to change the results. Even if the sample size were increased in practice, these estimators may still not perform well because FIA has a fixed sampling intensity of approximately one plot

Table 6. Results of the fourth simulation. Rows are arranged in ascending order according to average RMSE within each size class.

Size class	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Sawtimber	TS1(SNR = *opt*)	0.0165	0.960	10
	MA	0.0166	0.977	10
	TS1(SNR = 1)	0.0185	0.198	0
	LL	0.0209	0.051	0
	TS2(SNR = 0,*opt*)	0.0209	0.050	0
	TS2(SNR = 0,1)	0.0220	0.029	0
	SA	0.0272	0.004	0
Poletimber	TS1(SNR = *opt*)	0.0213	0.866	7
	MA	0.0219	0.919	7
	TS1(SNR = 1)	0.0236	0.210	0
	LL	0.0267	0.069	0
	TS2(SNR = 0,*opt*)	0.0267	0.068	0
	TS2(SNR = 0,1)	0.0279	0.043	0
	SA	0.0344	0.010	0
Sapling	TS1(SNR = *opt*)	0.0091	0.558	3
	MA	0.0095	0.729	3
	TS1(SNR = 1)	0.0110	0.109	0
	LL	0.0126	0.081	0
	TS2(SNR = 0,*opt*)	0.0126	0.077	0
	TS2(SNR = 0,1)	0.0132	0.045	0
	SA	0.0163	0.010	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

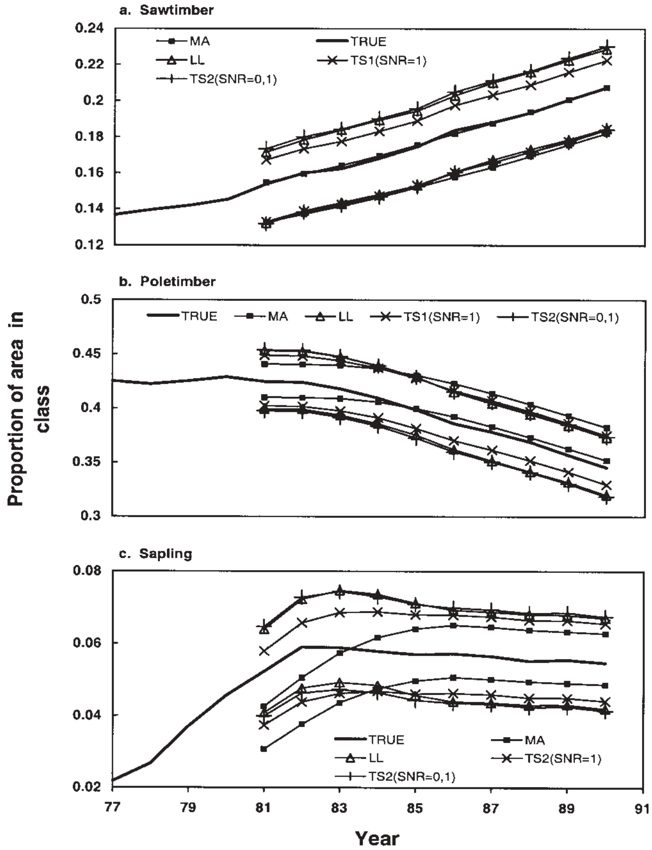

Figure 6. Estimator performance in the fourth simulation (condition class variables). The points lines represent a 1 standard deviation band for a given estimator over 5,000 simulated samples. The *TS*1(*SNR* = *opt*) and *TS*2(*SNR* = 0,*opt*) were excluded because they were virtually identical the the *MA* and *LL* estimators.

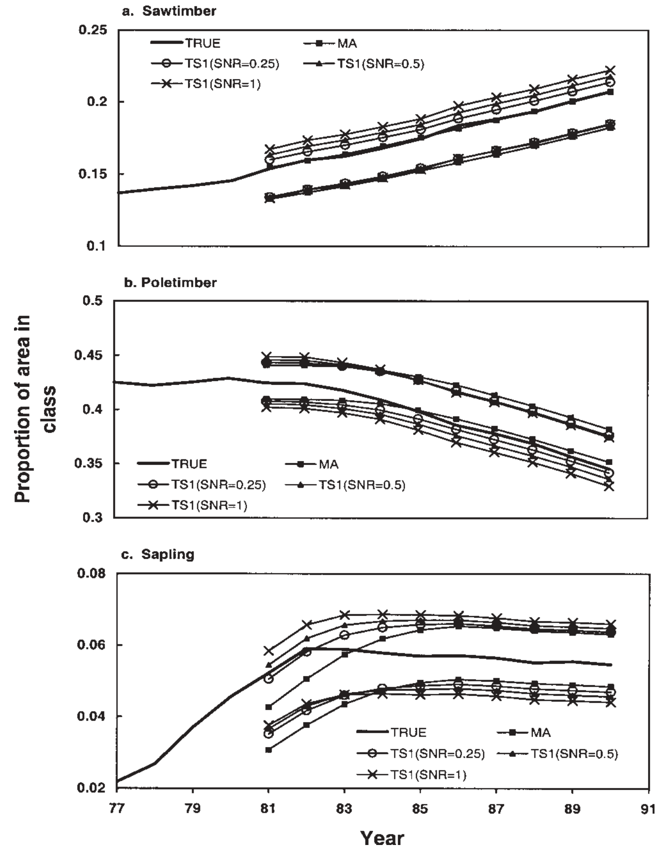

Figure 7. Performance of the *TS*1 estimators in the fifth simulation (condition class variables). The lines represent a 1 standard deviation band for a given estimator over 5,000 simulated samples.

Table 7. Results of the fifth simulation (TS1 estimators). Rows are arranged in ascending order according to average RMSE within each size class.

Size class	Estimator	Ave. RMSE	Ave. BR*	BR > 0.5†
Sawtimber	TS1(SNR = 0.25)	0.0155	0.527	7
	MA	0.0165	0.987	10
	TS1(SNR = 0.5)	0.0165	0.340	0
	TS1(SNR = 1)	0.0184	0.195	0
	LL	0.0207	0.054	0
	TS2(SNR = 0,0.25)	0.0210	0.046	0
	TS2(SNR = 0,0.5)	0.0212	0.040	0
	TS2(SNR = 0,1)	0.0218	0.031	0
	SA	0.0270	0.007	0
Poletimber	TS1(SNR = 0.25)	0.0203	0.512	6
	TS1(SNR = 0.5)	0.0214	0.341	3
	MA	0.0219	0.907	7
	TS1(SNR = 1)	0.0236	0.203	0
	LL	0.0268	0.066	0
	TS2(SNR = 0,0.25)	0.0269	0.057	0
	TS2(SNR = 0,0.5)	0.0272	0.050	0
	TS2(SNR = 0,1)	0.0279	0.040	0
	SA	0.0344	0.012	0
Sapling	TS1(SNR = 0.25)	0.0093	0.342	2
	MA	0.0096	0.725	3
	TS1(SNR = 0.5)	0.0100	0.209	1
	TS1(SNR = 1)	0.0111	0.113	0
	LL	0.0127	0.081	0
	TS2(SNR = 0,0.25)	0.0128	0.068	0
	TS2(SNR = 0,0.5)	0.0129	0.059	0
	TS2(SNR = 0,1)	0.0133	0.046	0
	SA	0.0164	0.011	0

* The true BR for the SA estimator is 0, but, the simulated BR is presented for comparison.
† This column represents the number of years in which BR > 0.5.

per 2,400 ha. Thus, increasing the sample size also means that larger areas must be considered, and it is possible that the spatial variance will increase with this increase in area.

Of the estimators studied, the *TS*1 class seems to have enough flexibility to handle estimation in a multiresource survey. The model parameters are easily interpreted with regard to variance structure, making it possible to be more objective in choosing constant values for this parameter. In addition, the *MA* estimator worked well in this simulation study as long as spatial variation was large. Therefore, we suggest using the *TS*1 estimators to easily incorporate previous years, data into the current year's estimates. We would like to caution, however, that these results are based on data from a specific region of Minnesota, which implies that the results may not be identically replicated in practice when applied to other forest types or natural resources. The main result that one should draw from this research is that the moving average should not be so quickly dismissed due to bias problems. The real problem is estimator mean square error.

Literature Cited

Breidt, F.J. 1999. Estimation systems for rolling samples. Forest inventory and monitoring biometricians workshop. Unpublished notes available from the author.

Brockwell, P.J., and R.A. Davis. 1987. Time series: Theory and methods. Springer-Verlag, New York. 577 p.

Czaplewski, R.L. 1995. Continuous adaptive monitoring of status and trends in ecosystem conditions. P. 80–85 *in* Proc. of the 1995 Soc. of Am. For. Conv. Soc. of Am. For., Bethesda, MD.

Czaplewski, R.L. 1999. Multistage remote sensing toward an annual inventory. J. For. 97:44–48.

Eriksson, M. 2001a. An examination of the mapped-plot design using an associative paradigm. Manuscript available from the authors at http://www.forestry.tamu.edu/papers/mapped-plots.pdf.

Eriksson, M. 2001b. Randomization-based weights in an annualized survey. Manuscript available from the authors at http://www.forestry.tamu.edu/papers/opt-wts.pdf.

Eriksson, M. 1995. Design-based approaches to horizontal-point sampling. For. Sci. 41:880–907.

Gillespie, A.J.R. 1999. Rationale for a national annual inventory program. J. For. 97:16–20.

Jakes, P.J. 1980. The fourth Minnesota forest inventory: Area. USDA For. Serv. Res. Bull. NC-54.

Leatherberry, E.C., J.S. Spencer, T.L. Schmidt, and M.R. Carroll. 1995. An analysis of Minnesota's fifth forest resources inventory, 1990. USDA For. Serv. Res. Bull. NC-165.

McRoberts, E.R. 1999. Joint annual forest inventory and monitoring system: The north central perspective. J. For. 97:27–31.

McRoberts, E.R., and M.H. Hansen. 1999. Annual forest inventories for the north central region of the United States. J. Agric. Biol. Environ. Stat. 4:361–371.

Neter, J., M.H. Kutner, C.J. Nachsheim, and W. Wasserman. 1989. Applied linear statistical models. Irwin, New York. 1,408 p.

Reams, G.A., and P.C. Van Deusen. 1999. The southern annual forest inventory system. J. Agric. Biol. Environ. Stat. 4:346–360.

Ripley, B.D. 1981. Spatial statistics. Wiley, New York. 272 p.

Robinson, A.P., D.C. Hamlin, and S.E. Fairweather. 1999. Improving forest inventories: Three ways to incorporate auxiliary information. J. For. 44:38–42.

Roesch, F.A., and G.A. Reams. 1999. Analytical alternatives for an annual inventory system. J. For. 97:33–37.

Sarndal, E.C., B. Swensson, and J. Wretman. 1992. Model assisted survey sampling. Springer, New York. 694 p.

Schreuder, H.T., T.G. Gregoire, and G.B. Wood. 1993. Sampling methods for a multiresource inventory. Wiley, New York. 464 p.

Van Deusen, P.C., S.P. Prisley, and A.A. Lucier. 1999. Adopting an annual inventory system: User perspectives. J. For. 97:11–14.