Canadian Journal of Forest Research

Evaluation of mapped-plot variance estimators across a range of partial nonresponse in a post-stratified national forest inventory

Journal:	<i>Canadian Journal of Forest Research</i>
Manuscript ID	cjfr-2021-0159.R1
Manuscript Type:	Note
Date Submitted by the Author:	27-Jul-2021
Complete List of Authors:	Westfall, James; USDA Forest Service Lister, Andrew; Northern Research Station, Forest Inventory and Analysis Scott, Charles; U.S Forest Service (retired)
Keyword:	nonforest, ratio estimate, bias, correlation, ratio-to-size
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Evaluation of mapped-plot variance estimators across a range of partial nonresponse in a post-stratified national forest inventory

James A. Westfall^{1*}, Andrew J. Lister¹, Charles T. Scott²

¹ U.S. Forest Service Northern Research Station, 3460 Industrial Drive, York, PA, 17402 U.S.A

² U.S. Forest Service (retired), West Chester, PA, 19382 U.S.A

* james.westfall@usda.gov

Draft

Abstract

When conducting a forest inventory, sometimes portions of plots cannot be measured due to inaccessibility. Two primary methods have been presented to account for partial nonresponse in the estimation phase: 1) use a ratio-to-size estimator, or 2) apply an adjustment factor to all plot observations in proportion to the missing area. Both approaches provide identical estimates of the population mean, but the estimates of variance differ when partial nonresponse is present. Variance estimator performance was examined for a range of population forest area and partial nonresponse proportions in the sample. The ratio-to-size variance estimator performed unbiasedly with respect to simulation results, but the adjustment factor variance estimates were biased with the magnitude and direction dependent upon the forest area proportion and amount of partial nonresponse. The bias is relatively small when the partial nonresponse is small, which is often the case; however, the ratio-to-size method is preferred to ensure accurate variance estimation for a wide range of circumstances.

Keywords: nonforest, ratio estimate, bias, correlation

Introduction

Various factors affect the specification of field plot measurement protocols in forest inventories. For large-area inventories where many different land characteristics are likely to be encountered, a mapped-plot protocol is used to describe the amount and type of various conditions that may be present (Wakelin et al. 2016, Bechtold and Scott 2005, Schreuder and Gregoire 1995). In addition to the flexibility for mapping differing conditions, the mapped-plot protocol also allows for delineation of nonsampled portions of plots due to lack of access permission or hazardous

conditions. In general, there are two primary estimation approaches for handling the partial nonresponse in the estimation phase: 1) use a ratio-to-size method where the area actually sampled is used instead of the full sample plot area (Van Deusen 2004), or 2) applying an average adjustment to all plot observations based on the amount of partial nonresponse in the sample (Scott et al. 2005). Both approaches provide identical estimates of the population mean for a given sample, however empirical results for the estimated variance differ. Van Deusen (2005) used a simulated population to show the average adjustment method overestimated the variance proportionally to the amount of partial nonresponse. Velasco-Bautista et al. (2020) found similar results using an artificial mapped tree population based on national forest inventory (NFI) data in Mexico. These studies showed the two approaches to be equal when no partial nonresponse was present and nearly so when partial nonresponse was rarely encountered. Thus, use of the average adjustment method may have small practical impact for NFIs having little partial nonresponse.

Although it is not specifically stated in either the Van Deusen (2005) or Velasco-Bautista et al. (2020) studies, it appears that most (if not all) plot observations were > 0 . In the case of many NFIs, the population structure can differ substantially from this configuration, as the population definition includes all lands within the country borders and thus nonforest plots contribute a valid observation of zero for estimation of forest attributes. Thus, further investigation is warranted to evaluate the partial nonresponse estimation approaches in a typical NFI application. Specifically, the performance of the variance estimators needs to be broadly assessed across a range of forest land area proportions in the population and varying amounts of partial nonresponse that may be realized in a NFI sample to better understand the factors that affect estimation accuracy.

Materials and Methods

Data

The data used to evaluate estimator performance were collected during 2013 to 2019 across an area of approximately 9,755,000 hectares in the state of Pennsylvania, U.S.A. by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service. FIA employs a quasi-systematic sampling design with a sampling intensity of approximately 1 plot per 2,428 ha (5,937 acres) (Reams et al. 2005). The field data used in this study were collected using a 0.067 ha (0.166 acre) cluster plot design, where each plot is composed of four circular subplots having 7.32 m (24 ft) radius with one subplot located at plot center and the remaining three subplots centered at azimuths 120, 240, and 360 degrees and distance of 36.58 m (120 ft) from plot center (Figure 1; Bechtold and Scott 2005). In forested areas, trees having a diameter at breast height (DBH) ≥ 12.70 cm (5.0 in.) are measured and their locations relative to subplot center are recorded. Observed tree data are subsequently used in statistical models to estimate tree wood volume (Scott 1981). In this study, wood volume is the attribute of interest used to evaluate different estimation approaches.

In some cases, portions of plots are nonsampled due to hazardous situations or lack of access permission and the boundaries of these areas are mapped such that the actual sampled area proportion can be calculated (U.S. Forest Service 2018). The FIA program gains considerable efficiency via pre-screening of plot locations using high-resolution remotely sensed imagery, such that entirely nonforest plots are not field visited and by definition have no trees or associated wood volume. These ‘office’ plots are still part of the inventory sample and, unlike

field plots, are not subject to nonresponse. This study only used 2,328 fully sampled plots that were entirely forest or nonforest, of which 905 were forest (i.e., forest area $\approx$ 39% of population area).

It is common in many NFIs to increase the precision of estimates through the use of auxiliary data. In the FIA program, this is accomplished via post-stratification where the post-strata are constructed from the National Land Cover Dataset (NLCD) tree canopy cover map (Homer et al. 2015) consisting of 30-m pixel-based estimates of canopy cover proportion between 0 and 1 across the U.S. To be consistent with the NFI implementation in the study area, five strata were defined by map canopy cover classes of 0-5% (stratum 1), 6-20% (stratum 2), 21-50% (stratum 3), 51-80% (stratum 4), and 81-100% (stratum 5). Plots were assigned to strata based on the pixel value corresponding with plot center. The stratum weights (W_h) were determined as the proportion of pixels in the canopy cover class divided by total number of pixels in the population and the weights were considered known quantities due to the complete spatial coverage of the NLCD data.

Table 1. Stratum size (N_h), weight (W_h), and summary statistics for forest area and volume in a post-stratified population of FIA plots in Pennsylvania, U.S.A. Note that each stratum contains nonforest plots (forest area proportion = 0) and plots having volume = 0 due to being either entirely nonforest or where forestland exists but no trees larger than 12.70 cm dbh are present.

Stratum	N_h	W_h	Forest area proportion			Volume (m ³ /ha)		
			Min.	Mean	Max.	Min.	Mean	Max.
1	1204	0.375	0.00	0.01	1.00	0.00	0.33	293.86
2	174	0.085	0.00	0.13	1.00	0.00	12.09	436.02
3	84	0.055	0.00	0.57	1.00	0.00	66.18	369.58
4	286	0.190	0.00	0.90	1.00	0.00	135.61	427.62
5	580	0.295	0.00	0.97	1.00	0.00	158.19	410.90
Total	2328	1.000	0.00	0.39	1.00	0.00	59.53	436.02

Post-stratified estimation

Given the interest in estimator performance under varying amounts of partial nonresponse, the estimators to be studied include 1) the partial nonresponse correction method as implemented by the FIA program (denoted FIA approach; Scott et al. 2005), 2) the mapped-plot estimator (VD) presented by Van Deusen (2004), and 3) the ratio-to-size estimator (RS) as described in an NFI context by Korhonen and Salmensuu (2014) with further elucidation by Scott and Westfall (2021). A mean of ratios (MOR) approach is not considered due the potentially high bias (Salas and Gregoire 2010) and conceivable implementation issues (Zarnoch and Bechtold 2000). A key differentiation between the FIA approach and the others is the modification of the plot-level observations to account for the partial nonresponse using:

$$y_{hi} = \frac{y'_{hi}}{\bar{a}_h}$$

where y'_{hi} is the the observed value expressed on a per hectare basis from the accessible portion of plot i in stratum h , $\bar{a}_h = \frac{\sum_i^n a_{hi}}{n_h}$ is the mean proportion of accessible plot areas in stratum h , a_{hi} is the proportion of accessible area of plot i in stratum h ($0 \leq a_{hi} \leq 1$), and y_{hi} is the i^{th} plot observation in stratum h used in estimation to account for partial nonresponse. Plots that are wholly nonresponse ($a_{hi} = 0$) are not included in the calculation of $\bar{a}_h$ and are not used in

estimation calculations. This adjustment approach is used to avoid potential bias resulting from excluding plots that are only partially observed. Within stratum h , all sampled plot areas are adjusted equally regardless of whether partial nonresponse exists on a plot (Scott et al. 2005).

Implementation of VD and RS methods require no plot-level adjustments, thus:

$$y_{hi} = y'_{hi}$$

For FIA, $\bar{y}_h = \frac{\sum_i^{n_h} y_{hi}}{n_h}$ is the estimated mean for attribute y in stratum h and n_h is the sample size for

stratum h . The estimated mean under both VD and RS methods is $\bar{y}_h = \frac{\sum_i^{n_h} y_{hi}}{\sum_i^{n_h} a_{hi}}$. Given that $\sum_i^{n_h} a_{hi}$

$= n_h \bar{a}$, the estimated $\bar{y}_h$ is identical for all three methods. However, this is not the case for

estimated variances, where the FIA method uses $v_{FIA}(\bar{y}_h) = \frac{\sum_i^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h(n_h - 1)}$ to obtain the estimated

variance of the mean for attribute y in stratum h , where N_h is the total number of population units in stratum h . The variance estimators differ accordingly with the estimated variances for VD and

RS methods being, respectively $v_{VD}(\bar{y}_h) = \frac{\sum_i^{n_h} (y_{hi} - a_{hi}\bar{y}_h)^2}{\sum_i^{n_h} a_{hi}(\sum_i^{n_h} a_{hi} - 1)}$ and $v_{RS}(\bar{y}_h) = \frac{n_h \sum_i^{n_h} (y_{hi} - a_{hi}\bar{y}_h)^2}{n_h - 1 (\sum_i^{n_h} a_{hi})^2}$.

Although finite population correction factors are usually unnecessary in NFI applications due to small sampling fractions, inclusion in the preceding formulae ($1 - n_h/N_h$ for FIA and $1 -$

$\sum_i^{n_h} a_{hi}/N_h$ for VD and RS) may be needed when large sampling fractions are used and were

included in the simulations described below. The respective overall estimated mean and variance

(Cochran 1977) for all methods are $\bar{y} = \sum_h^H W_h \bar{y}_h$ and $v(\bar{y}) = \frac{1}{n}$

$\left[\sum_h^H W_h n_h v(\bar{y}_h) + \sum_h^H (1 - W_h) \frac{n_h}{n} v(\bar{y}_h) \right]$ where n is the total sample size.

Simulation

The performance of the estimators described above were evaluated for estimates of mean wood volume per hectare in the population. A Monte Carlo simulation was conducted where for each replication the following steps were taken:

- 1) Select a 12% sample in each stratum proportional to stratum population size N_h . This sampling intensity was chosen to attain a minimum of $n_h = 10$ in each stratum (Westfall et al. 2011). The resultant stratum sample sizes were $n_1 = 143$, $n_2 = 21$, $n_3 = 10$, $n_4 = 34$, and $n_5 = 69$.
- 2) On forested plots, randomly choose 0 – 4 subplots for partial nonresponse.
- 3) For each subplot selected for nonresponse, randomly choose an azimuth from 0 – 359 degrees. To this azimuth, add the appropriate number of degrees for the level of nonresponse being investigated (30 – 330 degrees in increments of 30) (Figure 1).
- 4) Calculate the plot volume excluding the volume of any trees located within the nonresponse zone. Nonforest plots have zero volume by definition.

Two thousand replications were conducted for each partial nonresponse proportion class and the estimated means and variances for each method were calculated for comparative purposes.

Figure 1. FIA plot design and example of randomly generated 90-degree sections of partial nonresponse occurring on two subplots (shaded).

A broader simulation study was also conducted to examine a wider range of forestland area and partial nonresponse scenarios. This simulation used a synthetic population of 500 FIA plots where the forestland area proportions were controlled to span from 0.01 to 1.00 and the average partial nonresponse proportion on forested plots ranged from 0.0 to 0.5. Based on the specified combination of forestland area proportion and accessible proportion class, the following steps were taken at each iteration: 1) plots were randomly classified as either forest or nonforest, 2) total tree volume on forested subplots was randomly generated from a uniform distribution (0, 50) and nonforest subplots were assigned an observation of zero volume, 3) a subset of forested subplots were randomly chosen to be inaccessible and subplot volume a missing value, and 4)

means and variances were calculated based on $v_{FIA}(\bar{y}_h)$ and $v_{RS}(\bar{y}_h)$. Upon completion of 10,000 iterations, the mean values of the respective variances were calculated to evaluate the difference between the two methods.

Results

Given that each method provides an identical and unbiased estimate of the mean, the primary focus is on the performance of the variance estimators. For the population being considered (39% forest land), nonforest plots being excluded from partial nonresponse, and a post-stratified estimation framework, it was found that the FIA approach resulted in a smaller estimated variance compared to $v_{VD}(\bar{y})$, and $v_{RS}(\bar{y})$ when the total accessible area proportion within the sample plots was greater than ~ 0.86 (Figure 2). The FIA approach estimates a higher variance when the accessible area proportion is less than ~ 0.86 and the difference between FIA and the other approaches increases with decreasing accessibility proportion. Understanding this phenomenon requires examination of variances arising from each stratum as this is the level at which the estimators are applied.

Figure 2. Ratios of estimated variances arising from $v_{FIA}(\bar{y})$, $v_{VD}(\bar{y})$, and $v_{RS}(\bar{y})$ at various levels of accessible area proportions. Differences between $v_{VD}(\bar{y})$, and $v_{RS}(\bar{y})$ are small and both perform well regardless of completion rate.

A primary difference among strata are minimum accessible proportions attainable given the proportion of forested plots present. For example, stratum 1 contains about 1% forest area and therefore only about 1% of the plots are subject to partial nonresponse. Thus, even at high partial nonresponse rates, the minimum accessible proportion will be near 0.99. As strata become increasingly forested (Table 1), the potential minimum accessible proportion decreases accordingly. This point is relevant because the methods perform identically when the accessible proportion is 1.0 and thus deviations between the two approaches will be small within strata where completion rates are slightly less than 1.0 (e.g., stratum 1). A second outcome was that, similar to the findings of Velasco-Bautista et al. (2020), $v_{RS}(\bar{y})$ and $v_{VD}(\bar{y})$ produce variance estimates that are nearly identical (Figure 2), however, $v_{VD}(\bar{y})$ becomes slightly larger as accessible proportion diminishes. A third factor of considerable importance is the Monte Carlo variances for each stratum and accessible proportion matched very closely with $v_{RS}(\bar{y}_h)$ and thus deviations from $v_{RS}(\bar{y}_h)$ are considered as either over- or under-estimates of the true variance. These differences varied by stratum and accessible proportion, with those due to $v_{FIA}(\bar{y})$ being of primary interest where under-estimation occurred in strata 1 – 4, with an initial slight underestimation shifting to considerable overestimation in stratum 5 (Figure 3).

Figure 3. Ratio of $v_{FIA}(\bar{y}_h)/v_{RS}(\bar{y}_h)$ by stratum and accessible proportion.

The results show that for strata composed of approximately 90% forestland or less, partial nonresponse produces $v_{FIA}(\bar{y}_h)$ estimates that are too small. When forestland area exceeds this threshold, a shift occurs whereby $v_{FIA}(\bar{y}_h)$ overestimates the actual variance, with the overestimation increasing with increasing amounts of partial nonresponse. In this study, stratum 2 (13% forest) and stratum 3 (57% forest) show an asymptotic effect of accessible proportion on the underestimation of $v_{FIA}(\bar{y}_h)$ and stratum 4 (90% forest) exhibits a reversal in bias direction occurring at the accessible proportion of 0.77 (Figure 3). Stratum 5 (97% forest) initially has a small negative bias for $v_{FIA}(\bar{y}_h)$, but shifts to a positive bias near the 0.82 accessible proportion, with this positive bias increasing as accessible proportion continues to decline. This latter outcome is consistent with Van Deusen (2004) and Velasco-Bautista (2020) results pertaining to 100% (or nearly so) forestland area. It is suspected the slight initial negative bias in $v_{FIA}(\bar{y}_h)$ in

stratum 5 is due to the presence of a few nonforest plots that produce a weak negative covariance between a_i and y_i .

The observed outcomes using the FIA plot population led to a broader assessment of variance estimation behavior over a systematic range of forest area proportion and accessible proportion scenarios. The outcome aligned with the findings based on the FIA plot population, particularly the interaction that occurs between the proportion of forest area and the accessible proportion that results in a shift from under- to over-estimation of variance using $v_{FIA}(\bar{y}_h)$ (Figure 4). The results also agree with the result using $v_{FIA}(\bar{y})$ (Figure 2) where overall accessible proportion lower than about 0.88 results in underestimation of the variance; with over-estimation occurring at accessible proportions higher than ~ 0.88 . This study confirms the findings of both Van Deusen (2005) and Velasco-Bautista et al. (2020) using $v_{FIA}(\bar{y}_h)$ for fully forested populations, and provides a more complete picture of variance estimator behavior across a wide range of population structures common to national forest inventory applications.

Figure 4. Ratio of $v_{FIA}(\bar{y}_h)/v_{RS}(\bar{y}_h)$ for various levels of forest area proportion and accessible proportion.

Discussion

Generally, the FIA approach does not consider the correlation between a_i and y_i , i.e., $\bar{a}_h$ is a constant for a given sample. However, it is often the case that this assumption does not hold. Consider stratum 1, which is only 1% forest and is therefore subject to only a small amount of partial nonresponse. This stratum is composed mostly of nonforest plot observations where $a_i = 1$ and $y_i = 0$, with the few forest plots having $a_i = 1$ and $y_i > 0$ (no partial nonresponse) or $a_i < 1$ and $y_i > 0$ (partial nonresponse). This type of data structure results in a negative correlation between a_i and y_i . This same phenomenon occurs in other strata having $a_i = 1$ and $y_i = 0$ plots (primarily nonforest plots but could occur with forest plots with no tree volume). Correlations < 0 are present for strata 1 - 3 at all the partial nonresponse levels evaluated and strata 4 and 5 only have

correlation > 0 for relatively large inaccessible proportions (Figure 5). As the strata tend to become more forested (e.g., strata 4 and 5) and as the accessible proportions become smaller, the nonforest phenomenon that promotes negative correlations between a_i and y_i is diminished and positive correlations are realized because as a_i increases then y_i also tends to increase.

Figure 5. Correlation between a_{hi} and y_{hi} by stratum and accessible proportion.

The circumstance that FIA does not account for any correlation between a_{hi} and y_{hi} contributes to the inaccurate estimation of the stratum variance. Expansion of the numerator from the $v_{RS}(\bar{y}_h)$ estimator into $\sum_{i=1}^n y_{hi}^2 - 2\bar{y}_h \sum_{i=1}^n y_{hi} a_{hi} + \bar{y}_h^2 \sum_{i=1}^n a_{hi}^2$ explicitly reveals the covariance component as $-2\bar{y}_h \sum_{i=1}^n y_{hi} a_{hi}$. This term increases the overall variance when a negative covariance is present, but $v_{FIA}(\bar{y}_h)$ does not account for the covariance and thus is too small. Likewise, a positive within-stratum covariance should produce a commensurate decrease in the variance, but $v_{FIA}(\bar{y}_h)$ then overestimates the variance in this situation.

Given that the $v_{RS}(\bar{y}_h)$ estimated variances aligned with the Monte Carlo variances across all strata and accessible proportions, the $v_{RS}(\bar{y}_h)$ approach appears to be the preferred method for accurately estimating variance in the presence of partial nonresponse. This study examined a wide range of stratum accessibility proportions, many of which are unlikely to be encountered in practical application. An examination of stratum accessibility proportions across 24 states within the Northern FIA region (www.nrs.fs.fed.us/fia) showed a highly skewed distribution resulting from a few strata with relatively low accessible proportions (Figure 6). However, from a practical standpoint the mean accessible proportion over nearly 20,000 strata constructed from 2003 - 2019 was 0.995 with over half of the strata having 1.0 accessibility proportion. Thus, bias in stratum variances due to partial nonresponse using $v_{FIA}(\bar{y}_h)$ are generally very small and have minimal practical impact in creating bias in overall variance estimates produced by the FIA program. However, the results indicate that using $v_{FIA}(\bar{y}_h)$ where large partial nonresponse proportions are present can produce considerably inaccurate estimations of variance.

Figure 6. Distribution of number of strata within accessibility proportion classes across 24 states in the Northern U.S. from 2003-2019.

277

278 It is noteworthy that $v_{RS}(\bar{y}_h)$ is equivalent to the variance of a traditional ratio estimate (Cochran
 279 1977, p. 155) under the conditions that the variance of the ratio is: a) calculated with y'_{hi} (or,
 280 equivalently, y_{hi} with $\bar{a}_h=1$), and b) calculated only for the subpopulation consisting of
 281 accessible land. In this case, the ratio of means estimator expresses the result in terms of per unit
 282 area of the subpopulation; in our example, this would be *per hectare of accessible land*. The ratio
 283 is calculated as

$$284 \quad \hat{R}_h = \frac{\bar{y}'_h}{\bar{x}'_h}$$

285 where $\bar{y}'_h$ is calculated as

$$286 \quad \bar{y}'_h = \frac{\sum_i^{n_h} y'_{hi}}{n_h},$$

287 and $\bar{x}'_h$ is the average proportion of the stratum that is accessible and x'_{hi} is the proportion of plot i
 288 that is accessible, a_{hi} . The average of x'_{hi} is calculated in the same manner as for y'_{hi} . The
 289 variance of $\hat{R}_h$ is calculated as

$$290 \quad v(\hat{R}_h) = \frac{1}{\bar{x}_h'^2} [v(\bar{y}'_h) + \hat{R}_h^2 v(\bar{x}'_h) - 2 \hat{R}_h cov(\bar{y}'_h, \bar{x}'_h)],$$

291 where $v(\bar{y}'_h)$ is calculated as

$$292 \quad v(\bar{y}'_h) = \frac{(1 - n_h/N_h) \sum_i^{n_h} (y'_{hi} - \bar{y}'_h)^2}{n_h(n_h - 1)},$$

293 and $v(\bar{x}'_h)$ is calculated likewise. The covariance is calculated as

$$294 \quad cov(\bar{y}'_h, \bar{x}'_h) = \frac{(1 - n_h/N_h) \sum_i^{n_h} (y'_{hi} - \bar{y}'_h)(x'_{hi} - \bar{x}'_h)}{n_h(n_h - 1)}.$$

295 It is evident that $v_{RS}(\bar{y}_h)$ can be rewritten as

$$v_{RS}(\bar{y}_h) = \frac{n_h (1 - \frac{\sum_i^{n_h} a_{hi}}{N_h}) (\sum_i^{n_h} y_{hi}^2 - 2\bar{y}_h \sum_i^{n_h} y_{hi} a_{hi} + \bar{y}_h^2 \sum_i^{n_h} a_{hi}^2)}{n_h - 1 (\sum_i^{n_h} a_{hi})^2},$$

which is equivalent to Cochran's alternate expression of the traditional ratio estimate (Cochran 1977, p. 155, eq. 6.10), when subpopulation estimation is involved, i.e., when summations are done using y'_{hi} and x'_{hi} , which corresponds with the subpopulation of accessible land only and substituting $\sum_i^{n_h} a_{hi}$ for n_h .

Identifying this equivalency helps reveal the nature of a key difference between $v_{FIA}(\bar{y}_h)$ and $v_{RS}(\bar{y}_h)$, i.e., the way the partial plot correction ($\bar{a}_h$) is applied at the level of y_{hi} . This impacts the variance in the ways described above. In other words, $v_{FIA}(\bar{y}_h)$ applies $\bar{a}_h$ to each observation in the sample, as a method of accounting for missing portions of plots to obtain an unbiased estimate of the mean for the entire population. However, this approach has undesirable properties in the context of variance estimation. Using $v_{RS}(\bar{y}_h)$ or $v(\hat{R}_h)$ implicitly narrows the scope of inference to the subpopulation of accessible land, and relies on the data consumer to evaluate the validity of the assumption that non-sampled portions of the population are similar in status to the sampled portions. This assumption of similarity would be necessary to extend the results to the entire population that includes inaccessible areas. While there are pros and cons of both $v_{FIA}(\bar{y}_h)$ and $v_{RS}(\bar{y}_h)$ paradigms, the latter approach is more rigorous across a wide range of partial nonresponse outcomes.

Further study may be needed to extend the applicability to other forest inventory applications where nested fixed plot sizes or Bitterlich sampling methods are used to obtain plot observations. One possible approach when nested plot sizes are used is to scale the observations from smaller

318 plot sizes j to the largest plot size J . The plot attribute becomes $y_{hi} = \frac{\sum_j y_{hij} \sum_i^n a_{hij}}{\sum_i^n a_{hij}}$ where, within
 319 stratum h , y_{hij} is the observation for plot i on plot size j and a_{hij} is the plot size j area proportion
 320 measured on plot i . For variable size (Bitterlich) plots, the situation becomes more complex but
 321 in cases where full assessment of trees to be included in the sample is possible (but some
 322 physical tree locations cannot be accessed) a ratio based on accessible tally trees to total tally
 323 trees could be calculated on a basal area (or tree count) basis or if diameters of inaccessible trees
 324 can be adequately estimated then a similar ratio could be obtained based on sums of individual
 325 tree inclusion zone areas. When inaccessible tally trees cannot be assessed for sample inclusion,
 326 other methods would need to be devised. The performance of any given approach should be
 327 tested for validity prior to actual implementation in operational inventories.

328

329 **Conclusion**

330 This study extended the analytical realm of partial nonresponse scenarios to pertain more broadly
 331 to large-area forest inventories where more heterogeneous populations are often encountered.
 332 The results showed how interactions between forestland area proportion and sample plot
 333 accessibility proportion affect correlations between a_i and y_i , and the resultant effects when using
 334 the $v_{FIA}(\bar{y}_h)$ estimator. Bias in the overall variance estimate $v_{FIA}(\bar{y})$ depends both on the
 335 magnitude and direction of the stratum biases, as well as the stratum weights. This study did not
 336 evaluate directly the role of stratum weights in over- or under-estimation of the variance by v_{FIA}
 337 $(\bar{y})$; however, broad generalizations can be made, e.g., heavily forested populations will likely
 338 have larger weights in the high canopy cover strata (i.e., strata 4 and 5) and thus $v_{FIA}(\bar{y})$ will be
 339 more likely to over-estimate the overall variance, whereas populations with little forest would be

dominated by low canopy cover strata (i.e., strata 1 and 2) and presumably under-estimation would result.

Practitioners currently using a sample observation adjustment factor to account for partial nonresponse should evaluate their particular forest inventory circumstances. It was shown that $v_{FIA}(\bar{y})$ performs very similarly to $v_{RS}(\bar{y})$ when partial nonresponse is small (and identically when partial nonresponse is absent). Forest inventories such as those conducted on industry- or state-owned lands where entire plot access is essentially certain can use $v_{FIA}(\bar{y})$ with little concern. However, other types of inventories where plot access is dependent on various external factors that may result in nontrivial partial nonresponse should consider using $v_{RS}(\bar{y})$ to ensure accurate variance estimation across a range of potential nonresponse outcomes.

References

- Bechtold, W.A., and Scott, C.T. 2005. The forest inventory and analysis plot design. USDA Forest Service Gen. Tech. Rep. SRS-102, Southern Research Station, Asheville, NC. pp. 37-52.
- Cochran, W.G. 1977. Sampling Techniques, 3rd ed. John Wiley & Sons.
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. Photogramm. Eng. Remote Sens. **81**(5): 345-354.

363 Korhonen, K.T., and Salmensuu, O. 2014. Formulas for estimators and their variances in NFI.
364 Revised by Scott, C.T. An internal paper for the Forest and Agricultural Organization of the
365 United Nations. [https://www.dropbox.com/s/lhtpnom97p3a0gt/Formulas_KKor_2015-09-](https://www.dropbox.com/s/lhtpnom97p3a0gt/Formulas_KKor_2015-09-08%20Chip%20edits.docx?dl=0)
366 [08%20Chip%20edits.docx?dl=0](https://www.dropbox.com/s/lhtpnom97p3a0gt/Formulas_KKor_2015-09-08%20Chip%20edits.docx?dl=0).
367
368 Reams, G.A., Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A., and Moisen, G.G.
369 2005. The forest inventory and analysis sampling frame. USDA Forest Service Gen. Tech. Rep.
370 SRS-80, Southern Research Station, Asheville, NC. pp. 21-36.
371
372 Salas, C., and Gregoire, T.G. 2010. Statistical analysis of ratio estimators and their estimators of
373 variances when the auxiliary variate is measured with error. *Eur. J. For. Res.* 129(5):847-861.
374
375 Schreuder, H.T., and Gregoire, T.G. 1995. Estimating condition areas and boundary lengths in
376 sample plots with mapped designs. *Can. J. For. Res.* 25:1548–1553.
377
378 Scott, C.T. 1981. Northeastern forest survey revised cubic-foot volume equations. USDA Forest
379 Service Res. Note NE-304, Northeastern Forest Experiment Station, Broomall, PA. 3 p.
380
381 Scott, C.T., and Westfall, J.A. 2021. Ratio-to-Size Estimation. *In* Sampling and estimation
382 documentation for the enhanced Forest Inventory and Analysis program: 2020. Edited by J.A.
383 Westfall, J.W. Coulston, G.G. Moisen, and H.-E. Andersen. USDA Forest Service Gen. Tech.
384 Rep. NRS-____, Northern Research Station, Madison, WI. (In press).
385

U.S. Forest Service. 2018. Forest Inventory and Analysis national core field guide. Vol. 1: Field data collection procedures for phase 2 plots, version 7.2, Northern Research Station edition.

Available online at <https://www.nrs.fs.fed.us/fia/data-collection/field-guides/ver7.2/FG%20NRS%207.2-Complete%20Document.pdf>.

Van Deusen, P.C. 2004. Forest inventory estimation with mapped plots. *Can. J. For. Res.* 34(2):493-497.

Van Deusen, P.C. 2005. Comparing mapped plot estimators. In: Proceedings of the sixth annual Forest Inventory and Analysis symposium. Edited by R.E. McRoberts, G.A. Reams, P.C. Van Deusen and W.H. McWilliams. USDA Forest Service Gen. Tech. Rep. WO-70. Washington Office, Washington, DC. pp. 93-97.

Velasco-Bautista, E., de los Santos-Posadas, H., Ramírez-Maldonado, H., Rendón-Sánchez, G., Valdez-Lazalde, J.R., and Acosta-Mireles, M. 2020. Estimators for forest attributes on a simulated population of medium-sized tropical forest in southern Mexico. *Bosque* 41(3):307-320.

Wakelin, S., Searles, N., Paul, T., Kimberley, M. and Beets, P. 2016. New Zealand's national planted forest inventory as part of the Land Use and Carbon Analysis System. *New Zeal. J. For.* 61(2):23.

408 Westfall, J.A., Patterson, P.L., and Coulston, J.W. 2011. Post-stratified estimation: with-in strata
409 and total sample size recommendations. *Can. J. For. Res.* 41:1130-1139.
410
411 Zarnoch, S.J., and Bechtold, W.A. 2000. Estimating mapped-plot forest attributes with ratios of
412 means. *Can. J. For. Res.* 30(5):688-697.
413

Draft