Forestry: An International Journal of Forest Research, 2025, 1–9
https://doi.org/10.1093/forestry/cpaf032
Original Article

Double ratio estimation within a design-based nonresponse bias mitigation strategy

James A. Westfall[1,*] and Paul L. Patterson[2]

[1]U.S. Forest Service, Northern Research Station, 3460 Industrial Dr., York, PA 17402, United States
[2]U.S. Forest Service, Rocky Mountain Research Station, 240 W. Prospect Rd., Fort Collins, CO 80526, United States
[*]Corresponding author. U.S. Forest Service, Northern Research Station, York, PA 17402, United States. E-mail: james.westfall@usda.gov

Abstract

In the national forest inventory of the USA, there is an ongoing issue with sample plots that are either completely or partially unmeasured due to access issues such as hazardous conditions or lack of permission. The nonrandom nature of the nonresponse in the general population potentially creates a bias in the sample and resulting estimations. A potential mitigation option is the use of response homogeneity groups in combination with ratio-to-size estimators. This paper presents an extension of this approach where per-unit area estimates commonly useful to practitioners, e.g. number of trees per hectare, take the form of a double ratio (a ratio of ratios). Further complexity arises in the case of nested plot designs where attributes of interest need to be combined across plot sizes whose nonresponse rates may differ. The estimation details and associated formulae are presented and subsequently assessed using Monte Carlo simulation. The response homogeneity group ratio estimates appeared to be unbiased, but the approximate variances were substantially overestimated due to an overly large estimated covariance term. An alternative formulation of the estimated covariance produced overall variance estimates that aligned better with the simulation results but still were overestimated by ∼10% to 15%. The standard estimation approach resulted in the estimate being too large (0.3%) with an overall variance that was too small (−6.3%) due to substantially underestimated component variance and covariance terms. Thus, response homogeneity group estimation methods can produce unbiased per-unit area estimates in the presence of nonresponse, although it appears the estimated variance is somewhat conservative.

Keywords: approximate covariance; nested plots; forest inventory; ratio-to-size; sample bias

Introduction

Bias in calculated statistics due to sample unit nonresponse is a concern in many surveys. In large area forest inventories such as those conducted at a national scale National Forest Inventory (NFI), nonresponse is essentially inevitable due to factors such as hazardous conditions or lack of access permission. Furthermore, amounts of nonresponse can be highly variable in both spatial and temporal contexts (Patterson et al. 2012, Westfall et al. 2022). An important issue that is commonly encountered is that the nonresponse does not occur randomly with respect to population characteristics (Särndal and Lundström 2005). Various approaches have been investigated to account for potential nonresponse bias, including ignoring their occurrence (McRoberts 2003, Domke et al. 2014), post-stratification (Gormanson et al. 2023), response homogeneity groups (Goeking and Patterson 2013, Westfall 2022), calibration weighting (Fattorini et al. 2013, Corona et al. 2014, McCollum and Jacobs 2017), and imputation (Van Deusen 1997, Magnussen et al. 2017). Associated with these prospective remedies are a range of operational implementation nuances and complexities related to data management and estimation considerations. Ignoring the occurrence of nonresponse minimizes effects on data management (missing values) and estimation procedures (reduced sample size). However, in the likely situation of nonrandom nonresponse, this approach may also be the least robust to reducing nonresponse bias in the sample. Other methods require additional effort, such as tracking observed versus imputed sample units or conducting additional data processing to facilitate implementation of more complex estimation methods. Further issues may arise when considering forest inventories that conduct repeated measurements of permanent plots for detailed change information such as growth, removals, and mortality (Roesch 2007, Gschwantner et al. 2016).

Using NFI data from the state of New Mexico in the USA, Goeking and Patterson (2013) demonstrated the relative ease to implement response homogeneity groups (RHG; Särndal et al. 1992) into an existing post-stratified sampling design. In particular, creation of groups such as those separating sample plots into “office observation” and “field visit” were accomplished using readily available information in the database. Westfall (2022) conducted a similar study using NFI data from 10 states in the Northern United States. Across both efforts, apparent nonresponse biases of up to ∼4% in estimates of total forestland area were revealed. Thus, RHG may represent an appealing option for mitigating nonresponse bias while having only minor impacts on existing database structures and data processing algorithms. Nonetheless, the RHG approach does require more complex estimation procedures, particularly in the context of ratio-to-size estimators being used within RHG to address issues associated with partial plot nonresponse

Handling editor: Dr. Mark Ducey
Received 18 October 2024; **revised** 9 April 2025; **accepted** 16 May 2025

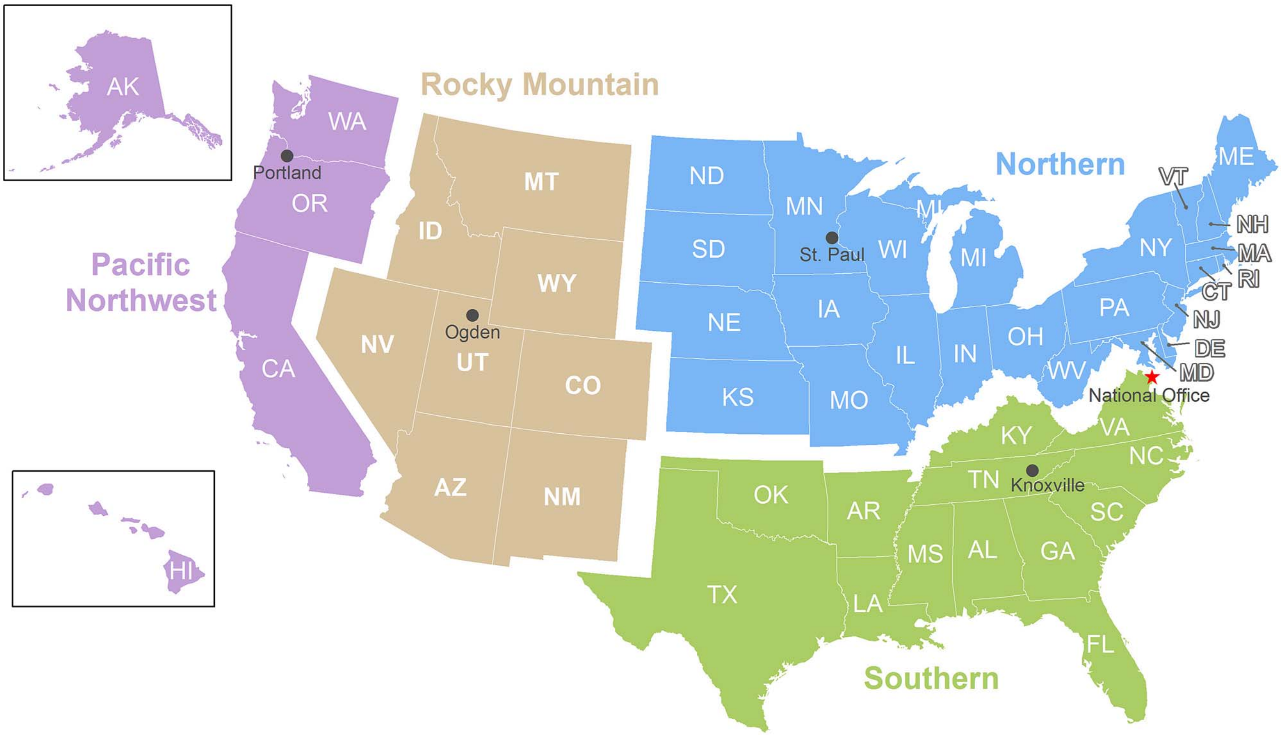

Figure 1. States within FIA functional region with regional and national office locations.

adjustment factors (Westfall 2022). Westfall (2022) presented the RHG formulae to calculate estimates of both totals and their variances for attributes such as forestland area and tree biomass, but often-desired estimates on a per-unit area basis (e.g. tree biomass per hectare of forestland) were not addressed. Thus, it is not clear how to accommodate fulfillment of these commonly generated forest resource statistics within the proposed RHG framework. The objectives of this paper are to (1) elucidate the formulae for calculating RHG per-unit area estimates and associated estimated variances, (2) present solutions for RHG estimation situations where attributes need to be combined across nested plot sizes, and (3) evaluate the performance of these estimators along with those from a standard post-stratified NFI estimation approach (Scott et al. 2005).

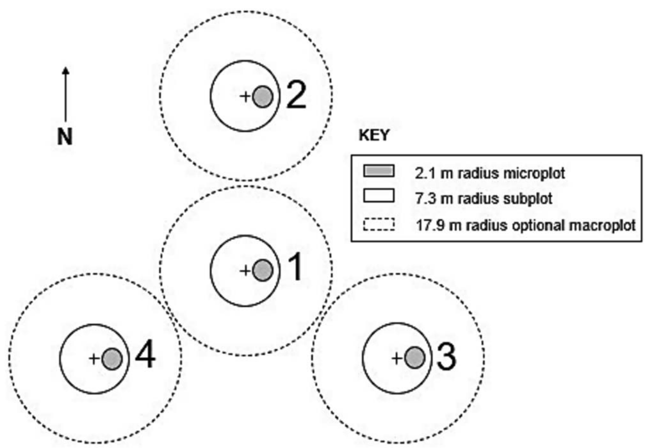

Figure 2. The FIA plot design.

Methods

Data

To provide context and examine the performance of the estimation methods, data collected for the national forest inventory of the USA by the Forest Inventory and Analysis (FIA) program were compiled. Specifically, sample plot data from the 24 states occurring in the Northern region (Fig. 1) were assembled using the most recent plot measurement up to 2019 to serve as a base population. The field data were collected on plots 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° and distance of 36.58 m (120 ft) from plot center (Fig. 2; Bechtold and Scott 2005). Subplot-based measurements within forested conditions include trees having diameter at breast height (dbh) $\geq$12.7 cm (5.0 in.) and various other site attributes such as forest type and stand size class. Each subplot contains a 2.1-m (6.8 ft) radius microplot with center offset 3.7 m (12 ft) at 90° azimuth. Sapling-size trees having 2.5 cm (1.0 in.) $\leq$ dbh $<$ 12.7 cm (5.0 in.) are recorded within the microplot.

An optional macroplot based at subplot center with a radius of 18.0 m (58.9 ft) may also be used. The purpose of the macroplot is to measure large trees exceeding a specified dbh threshold. The Northern FIA region does not measure macroplots; however, their inclusion in the description here facilitates a discussion point later in the manuscript.

A key aspect of the plot measurement protocol is mapping of differing conditions, e.g. between forested and nonforested portions of a plot. The mapping protocol allows for calculation of area proportions for each delineated condition. Of particular relevance to this study is the inclusion of mapping nonresponse portions of plots, e.g. nonresponse occurs as either the entire plot (complete) or a portion thereof (partial). In the standard FIA estimation process, complete nonresponse plots are not used and an adjustment factor is applied to plot observations based on the amount of partial nonresponse present in the sample (Scott et al. 2005).

Table 1. Summary of population size (N), numbers of trees (dbh $\geq$ 12.7 cm) and saplings (2.5 $\leq$ dbh < 12.7 cm), forest area, stratum weights (w_h), and RHG weights (w_{hg})

Stratum	RHG	N	No. trees	No. saplings	Forest area (ha)	w_h	w_{hg}
1	F	5484	5.6163E+05	2.0899E+06	2.3530E+03	0.6428	0.0782
1	O	64620	0	0	0		0.9218
2	F	4707	9.4411E+05	3.7862E+06	3.2548E+03	0.0678	0.6367
2	O	2686	0	0	0		0.3633
3	F	5230	1.5729E+06	6.0147E+06	4.3797E+03	0.0747	0.6422
3	O	2914	0	0	0		0.3578
4	F	23106	9.5850E+06	3.1248E+07	2.1875E+04	0.2148	0.9863
4	O	322	0	0	0		0.0137
All		109069	1.2664E+07	4.3139E+07	3.1863E+04		

The population of plots consisted of the FIA dataset described above but with only fully measured plots retained such that all attribute values are known. For this study, the number of trees and forestland area were chosen as the attributes of interest to facilitate estimation of the number of trees per hectare of forestland. In this case, the number of trees is specified as those having dbh $\geq$2.5 cm (1.0 in.) which entails using per-unit area expanded trees counts from both microplots and subplots. Forestland area estimation used the forest area proportions based on the subplot mapping.

Two elements common to many NFI are also applicable to these data: (1) aerial image-based assessments of forestland presence that are used to determine which plots are field visited (nonforest plots have zero forest and related attributes by definition, i.e. "office" plots) and (2) the use of post-stratification via auxiliary data to increase precision of estimates. In the context of this study, office plots are not subject to nonresponse because physically occupying the plot is unnecessary. The post-stratification mimics the canopy cover–based approach described in Gormanson et al. (2023) to develop $h = 1$ to H strata. Typically, post-stratum weights are determined as the proportion of the entire population belonging in class h. However, in this case the stratum weights (w_h) are based on plot proportions in each stratum to account for the unequal removal of nonresponse plots among strata. To facilitate analyses based on methods to be presented below, the office (O) or field (F) status of each plot is used to create RHG within each stratum h. The concept is to subdivide the post-strata developed for variance reduction purposes into groups of plots having similar response probabilities. An example of this strategy is to place office plots (100% probability of response) into one RHG and field plots (<100% probability of response) into a separate RHG. The RHG weights (w_{hg}) are based on the proportion of RHG plots in each stratum. Table 1 provides a description of relevant population statistics.

It should be noted that although the office and field RHG are used to illustrate the methods, RHG could be constructed from other or additional classifications (e.g. public vs private land ownership) where not all plots would be zero valued within a specific RHG such as for office plots. Thus, the methods as presented are generalizable to provide the broadest applicability to various NFIs whose circumstances pertaining to nonresponse may be different than those in the USA.

Estimation

To rationalize the estimation methods described below, pertinent details related to the sampling design and associated estimation paradigm are initially presented. The FIA sampling strategy consists of a quasi-systematic sample design with a post-stratified estimator that includes a compensation for partial nonresponse, and the traditional post-stratified variance estimator is used (Bechtold and Patterson 2005). It is assumed that the plots are a sample from a finite population and the post-stratified estimator applies to a simple random sample of the population. The quasi-systematic FIA sample is considered to be sufficiently similar to a simple random sample such that the properties of the post-stratified estimator are maintained. The post-stratified variance estimator is slightly conservative with respect to the true variance (Magnussen et al. 2020). Due to the availability of authentic forest inventory sample data, this study used the finite population approach instead of a continuous population scenario.

The primary motivation for this study was to address ratio estimation in the context of issues that arise due to unequal nonresponse within different nested plot sizes. When using ratio-to-size estimators and there is a need to combine attributes across nested plot sizes where the portions of inaccessible area (nonresponse) may differ between them, per-unit area estimates obtained from each plot size take the form of double ratios (Cochran 1977) and the overall population estimate arises from the sum of double ratio estimates. Considering the FIA subplot area for example, ratio-to-size estimators may be used to estimate the total number of trees having dbh $\geq$12.7 cm in the population ($\hat{y}$). The same method may be applied to the microplot for saplings having dbh 2.5 cm $\leq$ dbh < 12.7 cm to obtain an estimated population total ($\hat{y}'$). In pursuance of per-unit area estimates often preferred by foresters, assume that subplot forest area information is also used to estimate the total forestland in the population ($\hat{y}''$). The trees per hectare of forest estimates based on subplot and microplot areas respectively are defined as $\hat{R}_2 = \frac{\hat{y}}{\hat{y}''}$ and $\hat{R}_2' = \frac{\hat{y}'}{\hat{y}''}$. The "2" subscript on the left-hand side serves as a reminder of the double ratio construct. When the desired information is an estimate of the number of trees per hectare having dbh $\geq$2.5 cm, the sum of double ratio estimates $\hat{R}_2^{\Sigma}$ ensues:

$$\hat{R}_2^{\Sigma} = \frac{\hat{y}}{\hat{y}''} + \frac{\hat{y}'}{\hat{y}''} = \hat{R}_2 + \hat{R}_2' \tag{1}$$

Calculations to estimate the components of $\hat{R}_2$ and $\hat{R}_2'$ are relatively straightforward (Westfall 2022), and their variances $v\left(\hat{R}_2\right)$ and $v\left(\hat{R}_2'\right)$ can be estimated following the formulae given in Cochran (1977). The estimated variance of the double ratio sum $v\left(\hat{R}_2^{\Sigma}\right)$ can be generally constructed as

$$v\left(\hat{R}_2^{\Sigma}\right) = v\left(\hat{R}_2\right) + v\left(\hat{R}_2'\right) + 2\,\mathbf{cov}\left(\hat{R}_2, \hat{R}_2'\right) \tag{2}$$

Thus, the remaining task is to define the calculation procedures for $\mathbf{cov}\left(\hat{R}_2, \hat{R}_2'\right)$. However, in the spirit of providing a clear pathway to the intended outcome, considerable relevant background material is needed. As suggested above, pursuing this line of inquiry was predicated on minimizing potential empirical bias in estimation outcomes that may result from nested sample plots having unequal amounts of nonresponse. To broadly address the nonresponse issue, this study aims to employ RHG (Särndal et al. 1992, Goeking and Patterson 2013) in a post-stratified estimation context as described in Westfall (2022). Although the basic sampling design is considered to be a post-stratified simple random sample, implementation of RHG introduces a double sampling for post-stratification design within each of the post-strata. In this specific application, the first-stage sample is taken only at the plot locations and consists of office or field visit determinations that are used to estimate the RHG weights. The weights of the post-strata are determined from wall-to-wall remotely sensed information and are considered known values. As recommended by Westfall et al. (2022), the estimation process in Westfall (2022) uses ratio-to-size estimators for RGH means and variances to avoid bias in the estimated variance due to the correction factor used to account for partial nonresponse (Scott et al. 2005). The estimators for the RHG mean ($\bar{y}_{hg}$) and sample variance (s_{hg}^2) are restated here using the accessible proportion of the plot area instead of accessible plot area:

$$\bar{y}_{hg} = \frac{\sum_i^{m_{hg}} \tilde{y}_{hgi}}{\sum_i^{m_{hg}} p_{hgi}} \tag{3}$$

$$s_{hg}^2 = \frac{m_{hg}^2}{m_{hg}-1}\,\frac{\sum_i^{m_{hg}} \tilde{y}_{hgi}^2 - 2\bar{y}_{hg}\sum_i^{m_{hg}} \tilde{y}_{hgi}p_{hgi} + \bar{y}_{hg}^2\sum_i^{m_{hg}} p_{hgi}^2}{\left(\sum_i^{m_{hg}} p_{hgi}\right)^2} \tag{4}$$

Notationally, $\tilde{y}_{hgi}$ is the observed value for the attribute of interest from plot i in RHG g within post-stratum h, p_{hgi} is the accessible (response) area proportion from plot i in RHG g within post-stratum h where the p_{hgi} is assessed using the plot size $\tilde{y}_{hgi}$ is measured on, s_{hg} is the sample standard deviation in RHG g of post-stratum h, and m_{hg} is the sample size determined by the number of plots that were at least partially accessed. Estimation of post-strata means ($\bar{y}_h$) and variances $v\left(\bar{y}_h\right)$ proceeds in the typical manner; however, the estimated variance formula incurs a third term to account for estimation of the RHG weights from the sample and a fourth term reflects that nonresponse may vary from sample to sample.

$$\bar{y}_h = \sum_{g=1}^{G} w_{hg}\bar{y}_{hg} \tag{5}$$

$$v\left(\bar{y}_h\right) = \sum_{g=1}^{G}\left(\frac{n_{hg}'-1}{n_h'-1}\right)\frac{s_{hg}^2}{m_h} + \sum_{g=1}^{G}\left(\frac{n_{hg}'-1}{n_h'-1}\right)\frac{\left(1-w_{hg}\right)s_{hg}^2}{m_h^2 w_{hg}} + \frac{1}{n_h'-1}\sum_{g=1}^{G}\frac{n_{hg}'}{n_h'}\left(\bar{y}_{hg}-\bar{y}_h\right)^2 + \sum_{g=1}^{G}\left(\frac{n_{hg}'}{n_h'}\right)^2\frac{1-\sum_i^{m_{hg}} p_{hgi}/n_{hg}}{\sum_i^{m_{hg}} p_{hgi}}s_{hg}^2 \tag{6}$$

Note that the third term in (5) assumes the RHG weights are estimated; however, in some cases the RHG weights may be known (e.g. from a wall-to-wall map) and that term would be unnecessary. In the above, n notation is introduced to indicate sample sizes that would have occurred if there were no complete nonresponse plots. Also, n' notation is used to indicate the correspondence with the double sampling for post-stratification estimation paradigm. In keeping with previous notation, n_{hg}' is the sample size for RHG g in post-stratum h and n_h' is the sample size for post-stratum h in the absence of complete nonresponse. In this specific case where the first-stage sample occurs only at the plot locations, $n_{hg}' = n_{hg}$ and $n_h' = n_h$. The RHG weights are estimated from $w_{hg} = \frac{n_{hg}'}{n_h'}$. The value of m_h is consistent with the previous definition, i.e. the sum of the m_{hg} within post-stratum h. Subsequently, the estimated population total is calculated by combining the post-strata statistics:

$$\hat{y} = N\sum_{h=1}^{H} W_h\bar{y}_h \tag{7}$$

$$v\left(\hat{y}\right) = \frac{N^2}{m}\left[\sum_{h=1}^{H} W_h m_h v\left(\bar{y}_h\right) + \left(1-W_h\right)\frac{m_h}{m}v\left(\bar{y}_h\right)\right] \tag{8}$$

Note that the $h=1$ to H post-stratum weights (W_h) are considered to be known values obtained from wall-to-wall remotely sensed information, N is the population size, and m is the sum of the m_h over H strata.

This approach described above is relatively straightforward for estimating attribute means or totals as only a single response variable (e.g. $\tilde{y}_{hgi}$) must be considered. However, additional complexity is encountered in other scenarios such as estimating values on a per-unit area basis. For example, forest statistics such as number of trees per hectare of forestland are often more useful to forest managers than simply total number of trees. In a typical NFI scenario, this requires inclusion of a second response variable (e.g. to estimate forestland area) and a subsequent ratio estimate (i.e. total number of trees divided by forestland hectares). In the situation where ratio-to-size estimators such as (3) are used, the per-unit area calculation becomes a ratio of two ratio estimators (a "double ratio" per Cochran (1977)). As one might expect, estimation of the variance of a double ratio estimate is somewhat complicated and involves various covariance calculations (Cochran 1977). To facilitate elucidation of the formulae, begin by distinguishing between the numerator ($\hat{y}$) and denominator ($\hat{y}''$) attributes where e.g. the $\hat{y}$ may represent the estimated total number of trees with dbh $\geq$12.7 cm and the $\hat{y}''$ represents the estimated forestland area. Thus, the double ratio estimate ($\hat{R}_2$) is

$$\hat{R}_2 = \frac{\hat{y}}{\hat{y}''} \tag{9}$$

The "2" subscript on the left-hand side serves as a reminder of the double ratio construct. Cochran (1977) provides the estimated variance of $\hat{R}_2$ as

$$v\left(\hat{R}_2\right) = \left(\frac{\hat{y}}{\hat{y}''}\right)^2\left(\hat{c}_{\hat{y}\hat{y}} + \hat{c}_{\hat{y}''\hat{y}''} - 2\hat{c}_{\hat{y}\hat{y}''}\right) \tag{10}$$

where $\hat{c}_{\hat{y}\hat{y}} = \frac{v(\hat{y})}{\hat{y}^2}$, $\hat{c}_{\hat{y}''\hat{y}''} = \frac{v(\hat{y}'')}{\hat{y}''^2}$, and $\hat{c}_{\hat{y}\hat{y}''} = \frac{\mathbf{cov}(\hat{y},\hat{y}'')}{\hat{y}\hat{y}''}$. At this point, interest turns to estimating $\mathbf{cov}\left(\hat{y},\hat{y}''\right)$ as the values for all the other components have already been estimated. Calculations begin at the RHG level via estimation of $\mathbf{cov}\left(\bar{y}_{hg}, \bar{y}''_{hg}\right)$. Recall that

$\bar{y}_{hg}$ and yhg are both ratio estimates per (3). The estimated sample covariance formulation is based on an extension of the estimated sample variance formula given in (4) above:

$$\mathbf{cov}\left(\bar{y}_{hg}, \bar{y}''_{hg}\right) = \frac{m_{hg}^2}{(m_{hg}-1)} \tag{11}$$

$$\frac{\left(\begin{array}{c} \sum_i^{m_{hg}} \tilde{y}_{hgi}\widetilde{y''}_{hgi} - \bar{y}_{hg}\sum_i^{m_{hg}} \widetilde{y''}_{hgi}\dot{p}_{hgi} - \bar{y}''_{hg}\sum_i^{m_{hg}} \tilde{y}_{hgi}\ddot{p}_{hgi} \\ +\bar{y}_{hg}\bar{y}''_{hg}\sum_i^{m_{hg}} \dot{p}_{hgi}\ddot{p}_{hgi} \end{array}\right)}{\left(\sum_i^{m_{hg}} \dot{p}_{hgi}\right)\left(\sum_i^{m_{hg}} \ddot{p}_{hgi}\right)}$$

All the terms have been previously defined, but it is worthwhile to note that items with no superscript pertain to the numerator ratio-to-size estimator and those with the ″ superscript are related to the denominator ratio-to-size estimator. Furthermore, $\dot{p}_{hgi}$ and $\ddot{p}_{hgi}$ are more specific expressions of p_{hgi} designed to explicitly address different nested plot sizes with potentially differing proportions of sampled area. As depicted in (11), $\dot{p}_{hgi}$ refers to the plot size associated with the numerator ratio and $\ddot{p}_{hgi}$ pertains to the denominator ratio. $\dot{p}_{hgi} = \ddot{p}_{hgi}$in cases where both ratios are based on the same plot size. For all estimates, $\dot{p}_{hgi}$ and $\ddot{p}_{hgi}$ refer to the accessible area proportion of a plot (including nonforest area). Estimated post-strata covariances $\mathbf{cov}\left(\bar{y}_h, \bar{y}''_h\right)$ follow along the formulation of (6) as

$$\begin{aligned} \mathbf{cov}\left(\bar{y}_h, \bar{y}''_h\right) = & \sum_{g=1}^{G}\left(\frac{n'_{hg}-1}{n'_h-1}\right)\frac{\mathbf{cov}\left(\bar{y}_{hg}, \bar{y}''_{hg}\right)}{m_h} \\ & + \sum_{g=1}^{G}\left(\frac{n'_{hg}-1}{n'_h-1}\right)\frac{(1-w_{hg})\,\mathbf{cov}\left(\bar{y}_{hg}, \bar{y}''_{hg}\right)}{m_h^2 w_{hg}} \\ & + \frac{1}{n'_h-1}\sum_{g=1}^{G}\frac{n'_{hg}}{n'_h}\left(\bar{y}_{hg}-\bar{y}_h\right)\left(\bar{y}''_{hg}-\bar{y}''_h\right) \\ & + \sum_{g=1}^{G}\left(\frac{n'_{hg}}{n'_h}\right)^2 \frac{1-\sum_i^{m_{hg}}\dot{p}_{hgi}/n_{hg}}{\sum_i^{m_{hg}}\dot{p}_{hgi}}\,\mathbf{cov}\left(\bar{y}_{hg}, \bar{y}''_{hg}\right) \end{aligned} \tag{12}$$

Similarly, the final result for $\mathbf{cov}\left(\hat{y}, \hat{y}''\right)$ is estimated from

$$\mathbf{cov}\left(\hat{y}, \hat{y}''\right) = \frac{1}{m}\left[\sum_{h=1}^{H} W_h m_h\,\mathbf{cov}\left(\bar{y}_h, \bar{y}''_h\right) + (1-W_h)\frac{m_h}{m}\,\mathbf{cov}\left(\bar{y}_h, \bar{y}''_h\right)\right] \tag{13}$$

Recall now that the estimate of interest is the number of trees having dbh ≥2.5 cm per hectare of forestland, which requires combining information from two nested plot sizes. Due to the possibility of nonresponse amounts being different between the two plot sizes, it is necessary to calculate two double ratio estimates and sum them, i.e. number of saplings (2.5 ≤ dbh < 12.7 cm) per hectare forest plus number of trees (dbh ≥ 12.7 cm) per hectare forest. Clearly, the number of saplings estimate must arise from the microplot data, and it is only sensible to estimate number of trees from the larger subplot area. It is also practical to estimate forestland area based on the subplot as the larger plot size typically leads to a smaller estimated variance. In this context, a second double ratio may be defined as

$$\hat{R}'_2 = \frac{\hat{y}'}{\hat{y}''} \tag{14}$$

At this point, it may be considered that $\hat{y}$ represents the subplot-based estimate of number of trees, $\hat{y}'$ signifies the microplot-based estimate of number of saplings, and $\hat{y}''$ denotes the subplot-based estimate of forestland area. Thus, returning to the original concept as shown in (1) and (2), the final estimate and its estimated variance may be described by

$$\hat{R}_2^{\Sigma} = \frac{\hat{y}}{\hat{y}''} + \frac{\hat{y}'}{\hat{y}''} = \hat{R}_2 + \hat{R}'_2 \tag{15}$$

$$v\left(\hat{R}_2^{\Sigma}\right) = v\left(\hat{R}_2\right) + v\left(\hat{R}'_2\right) + 2\,\mathbf{cov}\left(\hat{R}_2, \hat{R}'_2\right) \tag{16}$$

As is commonly the case, the requisite values have been already calculated except for estimating $\mathbf{cov}\left(\hat{R}_2, \hat{R}'_2\right)$. Based on equation 6.95 in Cochran (1977), it is proposed that (Supplement 1)

$$\mathbf{cov}\left(\hat{R}_2, \hat{R}'_2\right) = \mathbf{cov}\left(\frac{\hat{y}}{\hat{y}''}, \frac{\hat{y}'}{\hat{y}''}\right) = \left(\frac{\hat{y}}{\hat{y}''}\right)\left(\frac{\hat{y}'}{\hat{y}''}\right)\left(\hat{c}_{\hat{y}\hat{y}'} + \hat{c}_{\hat{y}''\hat{y}''} - \hat{c}_{\hat{y}\hat{y}''} - \hat{c}_{\hat{y}'\hat{y}''}\right) \tag{17}$$

where $\hat{c}_{\hat{y}\hat{y}'} = \frac{\mathbf{cov}(\hat{y},\hat{y}')}{\hat{y}\hat{y}'}$, $\hat{c}_{\hat{y}''\hat{y}''} = \frac{\mathbf{cov}(\hat{y}'',\hat{y}'')}{\hat{y}''\hat{y}''} = \frac{v(\hat{y}'')}{\hat{y}''\hat{y}''}$, $\hat{c}_{\hat{y}\hat{y}''} = \frac{\mathbf{cov}(\hat{y},\hat{y}'')}{\hat{y}\hat{y}''}$, and $\hat{c}_{\hat{y}'\hat{y}''} = \frac{\mathbf{cov}(\hat{y}',\hat{y}'')}{\hat{y}'\hat{y}''}$. Note the term $\hat{c}_{\hat{y}''\hat{y}''}$ results from the common denominator in $\hat{R}_2$ and $\hat{R}'_2$. There may be situations where the denominators are different in the ratios and thus the term would be formulated differently, e.g. $\hat{c}_{\hat{y}'\hat{y}'''}$ if $\hat{R}'_2 = \frac{\hat{y}'}{\hat{y}'''}$. The terms $\hat{c}_{\hat{y}\hat{y}''}$ and $\hat{c}_{\hat{y}'\hat{y}''}$ are formulated based on equation 6.96 in Cochran (1977), specifically as $\hat{c}_{\hat{y}\hat{y}''} = \hat{c}_{\bar{y}_{hg}\bar{y}''_{hg}} + \hat{c}_{\bar{\dot{p}}_{hg}\bar{\ddot{p}}_{hg}} - \hat{c}_{\bar{y}_{hg}\bar{\ddot{p}}_{hg}} - \hat{c}_{\bar{y}''_{hg}\bar{\dot{p}}_{hg}}$ and $\hat{c}_{\hat{y}'\hat{y}''} = \hat{c}_{\bar{y}'_{hg}\bar{y}''_{hg}} + \hat{c}_{\bar{\dot{p}}_{hg}\bar{\ddot{p}}_{hg}} - \hat{c}_{\bar{y}'_{hg}\bar{\ddot{p}}_{hg}} - \hat{c}_{\bar{y}''_{hg}\bar{\dot{p}}_{hg}}$. As above, the $\hat{c}$ notation indicates an estimated covariance divided by the product of the respective estimates for terms on the right-hand side of the equality.

Initial investigations using the covariance estimator described above resulted in the postulation of two alternative approaches to also be evaluated. Option #1 was specifying an alternative $\hat{c}$ definition in the context of double ratios, where it was considered that the denominator be the double ratio estimate squared instead of the product of the two ratios, e.g. use $\hat{c}_{\hat{y}\hat{y}''} = \frac{\mathbf{cov}(\hat{y},\hat{y}'')}{\left(\frac{\hat{y}}{\hat{y}''}\right)^2}$ instead of $\hat{c}_{\hat{y}\hat{y}''} = \frac{\mathbf{cov}(\hat{y},\hat{y}'')}{\hat{y}\hat{y}''}$ and analogously for other terms. Option #2 builds upon the redefinition of $\hat{c}$ (option #1) and also stipulates a revised $\mathbf{cov}\left(\hat{R}_2, \hat{R}'_2\right) = \mathbf{cov}\left(\frac{\hat{y}}{\hat{y}''}, \frac{\hat{y}'}{\hat{y}''}\right) = \left(\frac{\hat{y}}{\hat{y}''}\right)\left(\frac{\hat{y}'}{\hat{y}''}\right)\left(\hat{c}_{\hat{y}\hat{y}'} + \hat{c}_{\hat{y}''\hat{y}''} - \hat{c}_{\hat{y}\hat{y}''} + \hat{c}_{\hat{y}'\hat{y}''}\right)$. This formulation is identical in form to (17) except the last term $\hat{c}_{\hat{y}'\hat{y}''}$ is now an addition instead of subtraction.

For comparative purposes, the standard FIA post-stratified estimation methodology was also used to generate estimates (Scott et al. 2005). Briefly, to account for plots having partial nonresponse, an adjustment factor is calculated for each post-stratum h and each nested plot size. This factor is used to adjust the attribute values based on their plot size for all observed plots (m_h) within the stratum. Classic post-stratified estimation then ensues using the adjusted values as observations. The standard FIA post-stratified estimator will be denoted by $\hat{R}^*$ and the variance estimator by $v\left(\hat{R}^*\right)$. Readers interested in this approach may consult Supplement 2 for the estimation formulae.

Simulation

Monte Carlo simulations were performed to assess the adequacy of the proposed estimators as well as the extent to which potential nonresponse bias is mitigated. The simulations were based on the FIA plot population described earlier, where at each iteration a sample of 5000 plots was chosen. Nonresponse was simulated by choosing randomly 10% of the field visited plots. Of those selected, 0.5%–1.5% were randomly chosen to have partial nonresponse and the remainder were complete nonresponse. An examination of FIA data showed that ~80% of plots having partial nonresponse had equal nonresponse proportions for subplots and microplots, mostly due to entire subplots being nonresponse and thus the nested microplot must be entirely nonresponse as well. As such, for 80% of the plots selected for partial nonresponse, a random variate from a U(0,1) distribution was chosen and used to represent the nonresponse proportion for both plot sizes. For the remaining 20% of partial nonresponse plots, the microplot nonresponse was determined from the U(0,1) random variate and the corresponding subplot nonresponse proportion was predicted from a simple linear model using the microplot nonresponse proportion as the predictor variable. The attribute values for each partial nonresponse plot were reduced proportionally to the sampled area remaining. Estimates of population totals and their variances were obtained using the estimation processes outlined above. Although not shown, finite population correction factors were used within the simulation for all variance and covariance estimates calculated at the RHG level. A Monte Carlo simulation was conducted with $q = 1$ to Q ($Q = 10\,000$) iterations and the mean value $\hat{\bar{R}}_2^{\Sigma}$ of the Q estimates of $\hat{R}_2^{\Sigma}$ were compared to the known population value to assess potential bias. Also, performance of the variance estimators was accomplished via assessment of the mean value $\bar{v}\left(\hat{R}_2^{\Sigma}\right)$ of $v\left(\hat{R}_2^{\Sigma}\right)$ against the variance of the Q estimates of $\hat{R}_2^{\Sigma}$ denoted as $v_{MC}\left(\hat{R}_2^{\Sigma}\right)$

$$\hat{\bar{R}}_2^{\Sigma} = \frac{\sum_q \hat{R}_2^{\Sigma}}{Q} \tag{18}$$

$$\bar{v}\left(\hat{R}_2^{\Sigma}\right) = \frac{\sum_q v\left(\hat{R}_2^{\Sigma}\right)}{Q} \tag{19}$$

$$v_{MC}\left(\hat{R}_2^{\Sigma}\right) = \frac{\sum_q \left(\hat{R}_2^{\Sigma} - \hat{\bar{R}}_2^{\Sigma}\right)^2}{Q\,(Q-1)} \tag{20}$$

Similar comparisons were made for estimates based on the standard FIA methods.

Results

Using the RHG estimation formulae, the population total and per-unit area estimates calculated from the $Q = 10\,000$ simulation replications were essentially unbiased when compared with the true population values (Table 2). With the exception of $\bar{v}\left(\hat{R}_2^{\Sigma}\right)$, the estimated variances for the population total and per-unit area estimates tended to be slightly large but never >5% (Table 2). Assessment of the variance estimate $v\left(\hat{R}_2^{\Sigma}\right)$ indicated poor agreement existed between $v_{MC}\left(\hat{R}_2^{\Sigma}\right) = 1749.42$ and $\bar{v}\left(\hat{R}_2^{\Sigma}\right) = 3765.52$ which overestimates the variance by ~115%. Further investigation revealed this outcome was partly due to associated conservative estimates for $v\left(\hat{R}_2\right) = 4.50\text{E}+10$ (vs $v_{MC}\left(\hat{R}_2\right) = 4.45\text{E}+10$) and $v\left(\hat{R}_2'\right) = 2.01\text{E}+12$ (vs $v_{MC}\left(\hat{R}_2'\right) = 1.93\text{E}+12$). Nonetheless, the result suggests the primary culprit for poor variance estimator performance was due to the estimated covariance between the double ratio estimates (17) that produced overly large values. Thus, all terms of (16) contributed to the overestimate of variance as judged against the Monte Carlo simulation values $v_{MC}\left(\hat{R}_2^{\Sigma}\right)$. Examination of Table 2 reveals an expected outcome that the largest variance contributor is from estimates of saplings on the microplot, i.e. smaller plot sizes often yield larger estimated variances than relatively larger plot sizes.

The most troubling outcome was the discrepancy between $v_{MC}\left(\hat{R}_2^{\Sigma}\right)$ and $\bar{v}\left(\hat{R}_2^{\Sigma}\right)$ that resulted from considerable overestimation of $\mathbf{cov}\left(\hat{R}_2, \hat{R}_2'\right)$. Although the formulation of $\mathbf{cov}\left(\hat{R}_2, \hat{R}_2'\right)$ appears to conform to statistical theory, Boes (1997) notes that "these approximate formulas can be extremely bad." The two alternative prescriptions for covariance estimation described earlier led to better results. Use of the option #1 formulation resulted in $\bar{v}\left(\hat{R}_2^{\Sigma}\right) = 2403.07$, which reduced the bias from 115% to 37%. Option #2 further reduced the bias with a result of $\bar{v}\left(\hat{R}_2^{\Sigma}\right) = 1908.76$, which was ~9% larger than $v_{MC}\left(\hat{R}_2^{\Sigma}\right) = 1749.42$. Given these alternative formulations may provide better approximations simply due to random chance or a specific population structure, evaluations of their performance were also conducted for three subpopulations constructed from the original data. These subpopulations consisted of aggregated data from several adjoining states (Fig. 1): (1) Ohio (OH), Pennsylvania (PA), and New York (NY); (2) Maine (ME), Vermont (VT), and New Hampshire (NH); and (3) Minnesota (MN), Wisconsin (WI), and Michigan (MI). Overall, the results were consistent with the initial finding in that using (17) to estimate the covariance produced the largest discrepancies between $\bar{v}\left(\hat{R}_2^{\Sigma}\right)$ and $v_{MC}\left(\hat{R}_2^{\Sigma}\right)$ (bias range 84%–167%), whereas the best results were obtained using option #2 (bias range 12%–16%).

Results from the standard FIA methods indicated negative biases in both forestland area (−1.74%) and numbers of stems (−1.43%) estimates due to the effect of nonresponse on the sample outcome (Table 2). However, a bias of 0.32% was realized in the ratio estimate as the direction and magnitude of estimate bias in the numerator and denominator are similar. The standard estimation also substantially underestimated the variances of both forestland area (−47.54%) and numbers of stems (−14.66%). The estimated covariance term in $v\left(\hat{R}^*\right)$ was also underestimated by −45.80%, which in combination with the other terms results in $v\left(\hat{R}^*\right)$ being −6.31% too small.

Discussion

Generally, the results showed the importance of explicitly accounting for effects of nonresponse when operating in a design-based framework. Under the assumption of equal response probabilities within RHG and adequate sample sizes, the RHG estimates appear to be effectively unbiased in the presence of nonresponse. In practice, the invocation of equal response probability for all field plots is a broad generalization and thus the proposed implementation does not strictly conform to RHG assumptions as presented in Särndal et al. (1992). However, the RHG method as implemented provides more accurate estimates compared to assuming all plots (including entirely nonforest "office" plots) have an identical response probability which is typical for standard estimation procedures. While most of the RHG estimated variances were slightly too large, the magnitude of the overestimation was relatively small. It is well established that obtaining unbiased estimators of variance for ratio estimates has been challenging. For example, Cochran (1977, Ch. 6) illustrates the difficulties in finding a suitable variance estimator for ratios that performs well across a range of applications. It is also noted that traditionally there has been purposeful emphasis on the

Table 2. Simulation results for estimates calculated from RHG and standard FIA methods

Estimation method	Attribute	Simulation mean estimate	Population value	Bias %	$v_{MC}(\hat{R}_2^{\Sigma})$	$\bar{v}(\hat{R}_2^{\Sigma})$	Bias %
RHG	Saplings ($\hat{y}'$)	4.312E+07	4.314E+07	-0.04%	1.93E+12	2.01E+12	3.97%
	Trees ($\hat{y}$)	1.266E+07	1.266E+07	0.01%	4.45E+10	4.50E+10	0.99%
	Forestland ($\hat{y}''$)	3.187E+04	3.186E+04	0.02%	7.31E+05	7.56E+05	3.36%
	Saplings/ha ($\hat{R}_2'$)	1353.16	1353.92	-0.06%	1.70E+03	1.78E+03	4.63%
	Trees/ha ($\hat{R}_2$)	397.42	397.44	-0.01%	3.46E+01	3.49E+01	0.88%
	All stems/ha ($\hat{R}_2^{\Sigma}$)	1750.58	1751.36	-0.04%	1.75E+03	3.77E+03	115.2%
Standard	Forestland ($\hat{y}^{*\prime\prime}$)	3.131E+04	3.186E+04	-1.74%	3.842E+13	2.015E+13	-47.54%
	All stems ($\hat{y}^*$)	5.500E+07	5.580E+07	-1.43%	6.457E+19	5.510E+19	-14.66%
	All stems/ha ($\hat{R}^*$)	1756.98	1751.36	0.32%	1823.078	1708.069	-6.31%

term "approximate" when referring to ratio estimator variance formulations (Hansen et al. 1953, Jessen 1978, Särndal et al. 1992).

The simulations also provided valuable insight into deficiencies of the standard estimation method in the presence of nonresponse. Estimates were too small for both total forestland area and number of stems. However, the magnitude of bias diminished for the ratio estimate as the biases in both the numerator and denominator tended to a canceling effect (Table 2). Särndal and Lundström (2005) comment that "… if an estimator is greatly biased, it is poor consolation that its variance is low." An analogous statement in the context of this result may be that it is of little solace to obtain an unbiased estimate that is derived from two biased estimates. Furthermore, the estimated variance formula (8) from the standard estimation produced substantially smaller values than obtained by calculating the variance of the Q simulation estimates for both estimates of total forestland area and total stems (Table 2). A limitation of this variance estimator is that the only sources of variability accounted for are the sampling variance and random sample sizes among post-strata. Thus, the estimated variance does not include any additional variation in estimates that arises due to a range of potential nonresponse outcomes. As may be expected, the estimated covariance was also considerably too small and a bias canceling effect was again present such that $\upsilon\left(\hat{R}^*\right)$ has less bias (−6.31%) than its constituent components. These biases are likely to increase in situations where the nonresponse rates exceed the 10% range used in this study (Westfall et al. 2022).

In this paper, the methods accounted for combining attributes that are measured on two nested plot sizes. Generally, the proposed estimation accounts for these different sizes, including estimated covariances among attributes occurring on the different plot sizes. It should be noted that when estimating these covariances, the last term in (12) requires the use of a single plot size. In this case, use of the largest plot size (subplot) is suggested and was used in this study. However, it is likely that little practical effect would be incurred if a different nested plot size were used as often the proportions of sampled areas are similar among nested plot sizes for clustered plot designs such as used here (Fig. 2). However, the validity of that assumption may be less robust when sample plots are of a singular nested design (Alberdi et al. 2017).

As suggested by Fig. 2, there may exist cases where three (or more) plot sizes must be considered. Conceivably, the methods presented could be extended to accommodate further nesting such as a sum of three double ratio estimates. This would require many more covariances to be estimated as well as careful construction of their formulation to ensure all the requisite terms have been accounted for. For example, if a third double ratio ($\hat{R}_2^{\Delta}$) was included, not only would $\mathbf{cov}\left(\hat{R}_2, \hat{R}_2'\right)$ be needed, but also $\mathbf{cov}\left(\hat{R}_2, \hat{R}_2^{\Delta}\right)$ and $\mathbf{cov}\left(\hat{R}_2', \hat{R}_2^{\Delta}\right)$. As was demonstrated in the analysis of two nested plot sizes, concern for the adequacy of the covariance approximations should not be overlooked.

Conclusion

The use of RHG within post-strata to reduce nonresponse bias can be highly effective, but also essentially creates a second level of post-stratification. When ratio-to-size estimators are used in this scenario, a relatively complex estimation situation arises when attributes measured on different nested plot sizes need to be combined and the amount of nonresponse differs among the nested plots. In this situation, development of the final population estimate required the sum of double ratio estimates. To the authors' knowledge, methods for calculating the variance of these sums have not appeared previously in the literature. The simulation results indicated that the approximate covariance estimator derived from statistical theory performed poorly when applied to the forest populations studied here and resulted in estimated variances that were too large. A proposed alternative covariance estimator provided considerable improvements in the estimated variances, although a modest amount of overprediction persisted. While the alternative approach may provide some practitioners with an acceptable variance estimate given the ratio estimation context, further study of both covariance formula specifications and their contributions to accurate variance calculations is recommended.

Although estimation pertaining to sums of double ratios is somewhat complex, implementation of the RHG methods is relatively user-friendly. It allows practitioners to remain in the well-understood realm of design-based estimation and is operationally efficient. The implementation investment is largely limited to a one-time effort to create the data processing and estimation algorithms. Furthermore, there is no disruption to the actual inventory data, which can be problematic from a temporal perspective when using some other nonresponse bias remedies such as imputation. It may also be worthwhile to consider additional factors that may impart nonrandom nonresponse outcomes such as public versus private land ownership. However, the curse of dimensionality will often limit the number of factors that may be accommodated while still maintaining adequate RHG sample sizes. Nonetheless, the RHG framework offers a flexible approach that can be adapted to nonresponse bias issues specific to a given forest inventory.

Acknowledgements

The authors are grateful to the Associate Editor and anonymous reviewers for their careful consideration of the manuscript content and comments for manuscript improvement.

Author contributions

James A. Westfall (Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft) and Paul L. Patterson (Conceptualization, Methodology, Validation, Writing—review & editing).

Supplementary data

Supplementary data are available at *Forestry* online.

Conflict of interest

None declared.

Funding

This work was supported by the U.S. Forest Service Northern Research Station and Rocky Mountain Research Station.

Data availability

The data underlying this article are available in the FIA DataMart at https://research.fs.usda.gov/products/dataandtools/tools/fia-datamart.

References

Alberdi I, Cañellas I, Bombín RV. The Spanish National Forest Inventory: history, development, challenges and perspectives. *Pesquisa Florestal Bras* 2017;**37**:361–8. https://doi.org/10.4336/2017.pfb.37.91.1337.

Bechtold WA, Patterson PL, (eds.). *The Enhanced Forest Inventory and Analysis Program - National Sampling Design and Estimation Procedures*. Gen. Tech. Rep. SRS-80. Asheville, NC, USA: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2005, 85.

Bechtold WA, Scott CT. *The Forest Inventory and Analysis Plot Design*. Gen. Tech. Rep. SRS-80. Asheville, NC, USA: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2005, 37–52.

Boes DC. *Unpublished Revision of Mood et al. (1974) Used in ST 520, Mathematical Statistics*. Colorado State University, 1997.

Cochran WG. *Sampling Techniques*. 3rd edition. New York, NY, USA: John Wiley & Sons, 1977, 428.

Corona P, Chirici G, Franceschi S. *et al.* Design-based treatment of missing data in forest inventories using canopy heights from aerial laser scanning. *Can J For Res* 2014;**44**:892–902. https://doi.org/10.1139/cjfr-2013-0521.

Domke GM, Woodall CW, Walters BF. *et al.* Strategies to compensate for the effects of nonresponse on forest carbon baseline estimates from the national forest inventory of the United States. *For Ecol Manage* 2014;**315**:112–20. https://doi.org/10.1016/j.foreco.2013.12.031.

Fattorini L, Franceschi S, Maffei D. Design-based treatment of unit nonresponse in environmental surveys using calibration weighting. *Biom J* 2013;**55**:925–43. https://doi.org/10.1002/bimj.201100262.

Goeking SA, Patterson PL. *Stratifying to Reduce Bias Caused by High Nonresponse Rates: A Case Study from New Mexico's Forest Inventory*. Res. Note RMRS-RN-59. Fort Collins, CO, USA: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2013.

Gormanson DD, Pugh SA, Barnett CJ. *et al. Statistics and Quality Assurance for the Northern Research Station Forest Inventory and Analysis Program, Version 3*. Gen. Tech. Rep. NRS-210. Madison, WI, USA: U.S. Department of Agriculture, Forest Service, Northern Research Station, 2023.

Gschwantner T, Lanz A, Vidal C. *et al.* Comparison of methods used in European National Forest Inventories for the estimation of volume increment: towards harmonisation. *Ann For Sci* 2016;**73**:807–21. https://doi.org/10.1007/s13595-016-0554-5.

Hansen MH, Hurwitz WN, Madow WG. *Sample Survey Methods and Theory. Vol. I. Methods and Applications*. New York, NY, USA: Wiley, 1953, 638.

Jessen RJ. *Statistical Survey Techniques*. New York, NY, USA: Wiley, 1978, 520.

Magnussen S, McRoberts RE, Breidenbach J. *et al.* Comparison of estimators of variance for forest inventories with systematic sampling-results from artificial populations. *For Ecos* 2020;**7**:1–19.

Magnussen S, Stinson G, Boudewyn P. Updating Canada's National Forest Inventory with multiple imputations of missing contemporary data. *For Chron* 2017;**93**:213–25. https://doi.org/10.5558/tfc2017-030.

McCollum JM, Jacobs DM. Calibration estimators in South Texas. In: *Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting. S.P. Healey, V.M. Berrett (comps.)*. Proc. RMRS-P-75. Fort Collins, CO, USA: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2017, 53–8.

McRoberts RE. Compensating for missing plot observations in forest inventory estimation. *Can J For Res* 2003;**33**:1990–7. https://doi.org/10.1139/x03-112.

Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics*. 3rd edition. New York, NY, USA: McGraw Hill, 1974, 480.

Patterson PL, Coulston JW, Roesch FA. *et al.* A primer for nonresponse in the US forest inventory and analysis program. *Env Mon Assess* 2012;**184**:1423–33. https://doi.org/10.1007/s10661-011-2051-5.

Roesch FA. Compatible estimators of the components of change for a rotating panel forest inventory design. *For Sci* 2007;**53**:50–61. https://doi.org/10.1093/forestscience/53.1.50.

Särndal C-E, Lundström S. *Estimation in Surveys with Nonresponse*. Chichester, England: John Wiley & Sons, 2005, 199. https://doi.org/10.1002/0470011351.

Särndal C-E, Swensson B, Wretman J. *Model Assisted Survey Sampling*. New York, NY, USA: Springer, 1992, 694. https://doi.org/10.1007/978-1-4612-4378-6.

Scott CT, Bechtold WA, Reams GA. *et al. Sample-Based Estimators Used by the Forest Inventory and Analysis National Information Management System*. Gen. Tech. Rep. SRS-80. Asheville, NC, USA: U.S. Department of Agriculture, Forest Service, Southern Research Station, 2005, 53–77.

Van Deusen PC. Annual forest inventory statistical concepts with emphasis on multiple imputation. *Can J For Res* 1997;**27**:379–84. https://doi.org/10.1139/x96-211.

Westfall JA. An estimation method to reduce complete and partial nonresponse bias in forest inventory. *Eur J For Res* 2022;**141**:901–7. https://doi.org/10.1007/s10342-022-01480-6.

Westfall JA, Schroeder TA, McCollum JM. *et al.* A spatial and temporal assessment of nonresponse in the national forest inventory of the U.S. *Env Mon Assess* 2022;**194**:530. https://doi.org/10.1007/s10661-022-10219-0.

Forestry: An International Journal of Forest Research, 2025, 1–9
https://doi.org/10.1093/forestry/cpaf032
Original Article