European Journal of Forest Research
https://doi.org/10.1007/s10342-022-01480-6

ORIGINAL PAPER

An estimation method to reduce complete and partial nonresponse bias in forest inventory

James A. Westfall[1]

Received: 23 August 2021 / Revised: 18 April 2022 / Accepted: 27 June 2022

Abstract
Survey practitioners commonly encounter various types of nonresponse and strive to implement methods that mitigate any resulting bias when reporting results. In national forest inventories (NFI), complete or partial nonresponse usually results from hazardous conditions or lack of plot access permission. While many factors may be related to nonresponse, the two primary factors in the NFI of the USA are public/private land ownership and office/field plot status. To ameliorate potential nonresponse bias, these factors should be accounted for in the estimation process. An estimation method is presented where response homogeneity groups (RHGs) account for differential nonresponse rates between forest/nonforest plots. In a post-stratified estimation context, ratio-to-size estimators are used in RHGs within post-strata to avoid potential bias in variance estimates arising from partial plot nonresponse. Combining RHGs within post-strata requires a complex variance estimator that includes four sources of uncertainty. Testing of the estimation method on a synthetic population showed the approach is essentially unbiased. Application to NFI data from 10 states in the USA consistently showed the RHG method produced state-level estimates of forestland area that were 0.1%–3.6% larger than the current post-stratified estimation procedure. It is suggested that these differences are indicative of the nonresponse bias present when plots having differential nonresponse rates are not accounted for.

Keywords Response probability · Double sampling · Post-stratification · Nonforest · Ownership

Introduction

Nonresponse is encountered in many types of surveys and considerable study has been devoted to the subject due to the potentially complex causal mechanisms and difficulty in avoiding sample bias. In national forest inventories (NFI), nonresponse is essentially unavoidable due to factors such as lack of land access permission and hazardous conditions (Patterson et al. 2012). Two types of nonresponse are often present: (1) the entire sample plot is inaccessible (complete nonresponse), or (2) a portion of the plot is inaccessible (partial nonresponse). Recognition of potential nonresponse bias had led to various attempts at mitigation, such as defining strata and/or subpopulations wherein the response probabilities for each sample plot are considered to be approximately equal and that the nonresponse plots have similar attributes as the response plots (Gormanson et al. 2018; Goeking and Patterson 2013). The other primary method for reducing nonresponse bias is to invoke an imputation process that assigns values where information is missing (Domke et al. 2014; Magnussen et al. 2017). However, difficulty often arises due to the complex array of possibly interrelated underlying factors that result in nonresponse, including private land ownership, hazardous conditions, land use, and possibly other characteristics such as forest type, stand volume, or tree density. In other words, it is very difficult to identify strata that fully characterize this complex situation. Unless the underlying mechanisms are properly identified and accounted for, the intended nonresponse bias mitigation outcomes are often not fully attained such that an unknown amount of bias is still present in empirical estimates.

Despite the difficulty in removing nonresponse bias entirely, situations where sources of nonresponse bias are known should be managed to the extent possible. In the case

Communicated by Lauri Mehtätalo.

James A. Westfall
james.westfall@usda.gov

[1] U.S. Forest Service, Northern Research Station, York, PA, USA

of the NFI in the USA conducted by the Forest Inventory and Analysis (FIA) program, two discernable factors are private land ownership and the subset of plots sent to the field for measurement. Nonresponse prevalently occurs on privately owned lands due to denial of access by the landowner, although on rare occasions access to public lands can be restricted. For efficiency, it is common in NFI to initially examine sample plot locations using high-resolution remotely sensed imagery to determine whether forest land is present and field measurement is required. For entirely nonforest plots, attributes such as forest area and associated tree volume are known to be zero without a field visit (office plot). As the plot area does not need to be physically visited, office plots are not subject to nonresponse. The issue of land ownership is addressed through creation of subpopulations intended to discretize public and private lands; however, differences in nonresponse rates between office and field plots have not been formally addressed in estimation processes. Response homogeneity groups (RHG) (Särndal 1992) can be used to separate office and field plots to mitigate nonresponse bias (Goeking and Patterson 2013). RHG are intended to group plots that have similar nonresponse probabilities. This study extends the analysis of Goeking and Patterson (2013) by presenting a variance estimator appropriate for the RHG design, describing a modified form of implementation, and examining potential nonresponse bias across a range of population characteristics and nonresponse rates to better understand the effects of ignoring the differences in office and field plot response probabilities.

Methods

Data

A synthetic population of 210,000 elements was created to test the estimation method. Each element is considered to represent a sample plot that could be selected when conducting an NFI. Two components of interest for each plot were office vs. field measurement status and presence of partial or complete nonresponse. Four strata were created to represent post-strata that are often employed as a variance reduction method. Typically, these strata would arise from maps derived from remotely sensed data with complete spatial coverage of the population such that stratum weights (W_h) are considered to be known without error. A generic response variable y_hi was generated for each plot i in stratum h where office plots were assigned $y_\mathrm{hi} = 0$ and field plots had y_hi values generated from a $\mathrm{N}(\mu_\mathrm{h}, \sigma_\mathrm{h})$ normal distribution with parameters for each stratum h shown in Table 1. While forest inventory response variables for field plots are generally not strictly normal in distribution, a cursory analysis suggested variables such as wood volume and biomass are

Table 1 Population size, known stratum weight, percent of office plots, and normal distribution parameters N(μh, σh) for four strata within the synthetic population

Stratum	N_h	W_h	Office plot %	μ_h	σ_h
1	60,000	0.286	90%	200	20
2	1,00,000	0.476	50%	400	40
3	40,000	0.190	10%	800	80
4	10,000	0.048	0%	1000	100
All	2,10,000	1.00	51%		

often unimodal and somewhat symmetric such that use of a normal distribution provides a reasonable approximation. These population structure and post-strata designations were designed to mimic a typical US NFI sample realization and subsequent estimation of cubic tree wood volume. Practitioners seeking to conduct similar analyses should develop the synthetic population in a manner that reflects likely inventory scenarios.

To account for differential nonresponse probabilities between office and field plots, response homogeneity groups (RHG; Särndal 1992, Sect. 15.6.2) were created within strata 1–3 to separate the field and office plots because office plots have zero probability of nonresponse. Stratum 4 does not use RHG due to being entirely comprised of field plots. The RHGs act as substrata where the weights are estimated from the sample as the proportion of plots in each group (Goeking and Patterson 2013):

$$w_\mathrm{hg} = {n'_\mathrm{hg}}\Big/{n'_\mathrm{h}} \tag{1}$$

where w_{hg} is the estimated weight for RHG g in stratum h, n'_hg is the number of plots in RHG g within stratum h, and n'_h is the number of plots in stratum h. The n' notation style is used to indicate this approach is considered equivalent to double sampling for post-stratification where the first phase sample only consists of the plot locations. In this case, sample sizes for RHGs and strata are $n_\mathrm{hg} = n'_\mathrm{hg}$ and $n_\mathrm{h} = n'_\mathrm{h}$, respectively.

Sampling and estimation

The analysis was conducted via a Monte Carlo simulation where at each of $r = 1, \ldots, 10{,}000$ iterations a sample of 400 plots was drawn randomly. Stratum membership was ignored in the selection process to mimic post-stratification where within-stratum sample sizes are a random outcome. Nonresponse was introduced via: (1) 10% of field plots were randomly chosen and designated as complete nonresponse, and (2) 3% of field plots were randomly chosen to have partial nonresponse where the inaccessible proportion was randomly generated from a (0,1) uniform distribution. The value of the response y_{hi} was reduced accordingly. It is important to distinguish here the difference between the number of plots selected to be sampled and the number of plots where observations were obtained. Referring back to (1), n notation indicates the number selected to be sampled, whereas m notation is now introduced to represent the actual number observed, i.e., m_{hg} is the number of realized plot observations in RHG g of stratum h. Entirely nonsampled plots are only used in estimation of the RHG weights (w_{hg}) and uncertainty due to nonresponse (as shown below).

The first step in the estimation process is to calculate means and sample variances for each RHG. Here and subsequently, variance estimators are presented without finite population correction due to the low sampling fraction common to most forest inventories. To avoid bias in the estimated variance due to partial nonresponse (Westfall et al. 2021; Van Deusen 2005), a ratio-to size estimator is used.

$$\bar{y}_{\mathrm{hg}} = \frac{\sum_i^{m_{\mathrm{hg}}} y_{\mathrm{hgi}}}{\sum_i^{m_{\mathrm{hg}}} a_{\mathrm{hgi}}} \tag{2}$$

$$s_{\mathrm{hg}}^2 = \frac{m_{\mathrm{hg}}^2}{m_{\mathrm{hg}} - 1} \frac{\sum_i^{m_{\mathrm{hg}}} y_{\mathrm{hgi}}^2 - 2\bar{y}_{\mathrm{hg}} \sum_i^{m_{\mathrm{hg}}} y_{\mathrm{hgi}} a_{\mathrm{hgi}} + \bar{y}_{\mathrm{hg}}^2 \sum_i^{m_{\mathrm{hg}}} a_{\mathrm{hgi}}^2}{\left(\sum_i^{m_{\mathrm{hg}}} a_{\mathrm{hgi}}\right)^2} \tag{3}$$

where $\bar{y}_{\mathrm{hg}}$ is the mean response for RHG g in stratum h, y_{hgi} is the response from the ith plot in RHG g for stratum h, a_{hgi} is the proportion of the plot that was accessible, and s_{hg}^2 is the sample variance of the response in RHG g for stratum h.

Subsequently, means and their variances for each of the h strata are needed and arise from combining the RHG within each stratum. There are several factors to be accounted for in the variance calculation beyond that due to sampling alone. When post-stratification is employed, the stratum sample sizes are random and an additional source of uncertainty (Cochran 1977, Sect. 5A.9) is indicated by the 2nd term in (5) below. Also, the weights for each RHG were estimated from the sample such that this source of uncertainty is accounted for via the 3rd term of (5). Essentially, this is a special case of double sampling for post-stratification where the first phase sample is only done at measured plot locations (Westfall et al. 2019). Finally, randomness of nonresponse represents additional uncertainty requiring the appropriate 4th term in the variance estimator (5) (Särndal 1992, Sect. 15.6.3). Typical construction of variance estimators that account for multiple sources includes a term for the sampling variance with further additive terms for other components. Westfall et al. (2019) developed a variance estimator that has additive terms for both double sampling and post-stratification uncertainty components. Building upon that work, the formulation of Särndal (1992, Eq. 15.6.15) serves as the basis for the fourth term corresponding to nonresponse uncertainty.

$$\bar{y}_h = \sum_{g=1}^{G} w_{\mathrm{hg}} \bar{y}_{\mathrm{hg}} \tag{4}$$

$$\begin{aligned} v(\bar{y}_h) = {} & \sum_{g=1}^{G} \left(\frac{n'_{\mathrm{hg}} - 1}{n'_{\mathrm{h}} - 1} \right) \frac{s_{\mathrm{hg}}^2}{m_{\mathrm{h}}} + \sum_{g=1}^{G} \left(\frac{n'_{\mathrm{hg}} - 1}{n'_{\mathrm{h}} - 1} \right) \frac{(1 - w_{\mathrm{hg}}) s_{\mathrm{hg}}^2}{m_{\mathrm{h}}^2 w_{\mathrm{hg}}} \\ & + \frac{1}{n'_{\mathrm{h}} - 1} \sum_{g=1}^{G} \frac{n'_{\mathrm{hg}}}{n'_{\mathrm{h}}} (\bar{y}_{\mathrm{hg}} - \bar{y}_{\mathrm{h}})^2 + \sum_{g=1}^{G} \left(\frac{n'_{\mathrm{hg}}}{n'_{\mathrm{h}}} \right)^2 \frac{1 - \sum_i^{m_{\mathrm{hg}}} a_{\mathrm{hgi}} / n_{\mathrm{hg}}}{\sum_i^{m_{\mathrm{hg}}} a_{\mathrm{hgi}}} s_{\mathrm{hg}}^2 \end{aligned} \tag{5}$$

It should be noted here that RHG may not be constructed within some strata due to small sample sizes or lack of both office and field plots, e.g., stratum 4 in the synthetic population (Table 1). In this situation, the above procedures can be implemented where the stratum is considered an RHG (G = 1). Note that in this case there is no double sampling involved in estimating the weight ($w_{hg} = 1$) such that the third term of (5) should not contribute to additional variance. This result is obtained because $\bar{y}_{\mathrm{hg}} = \bar{y}_{\mathrm{h}}$ and therefore $\bar{y}_{\mathrm{hg}} - \bar{y}_{\mathrm{h}} = 0$. Similarly, randomness in RHG sample size is not accounted for as the second term equals 0 when $w_{\mathrm{hg}} = 1$. This source of uncertainty is accounted for in the next estimation step described below.

The stratum means and variances can then be combined to estimate the population total and its variance (Scott et al. 2005):

$$\hat{y} = N \sum_{h=1}^{H} W_h \bar{y}_h \tag{6}$$

$$v(\hat{y}) = \frac{N^2}{m} \left[\sum_{h=1}^{H} W_h m_h v(\bar{y}_h) + (1 - W_h) \frac{m_h}{m} v(\bar{y}_h) \right] \tag{7}$$

where N = number of sample units in the population. If the generic response y_{hgi} is considered to be expressed on a per hectare basis, the area of the population in hectares can be used in place of N. The second term accounts for the random sample size for each stratum h that occurs due to post-stratification.

Evaluation of estimator performance was assessed by comparing the mean of the $R = 10{,}000$ population total estimates from the simulation with the known population mean $\mu = 59{,}884{,}824$. Similarly, the mean of the estimated variances $v(\hat{y})$ was compared the variance of the $R = 10{,}000$ population estimates.

Application of the estimation methods to actual NFI data was accomplished using FIA data from 10 states (New York, West Virginia, South Dakota, Iowa, Ohio, Maine, Illinois, Minnesota, Michigan, and Missouri) within the Northern FIA region. The states were chosen to represent a wide range of forest land area amounts and nonresponse rates. These data include indicators of field or office visit and for each plot a measurement of the proportion of the plot area being forest land. Plots are assigned spatially to subpopulations (termed estimation units) based on a public/private ownership map, with the specific intent to separate plots into public and private ownership due to nonresponse almost exclusive occurring on private lands, i.e., an effort has already been made to reduce nonresponse bias. Within estimation units, 1–5 strata are created using tree canopy cover classes with plots assigned to strata depending on their canopy cover from the map pixel associated with plot center location. Five strata are initially specified, but in some cases strata need to be collapsed to meet minimum within-stratum sample size requirements (usually 10). Two RHGs within each stratum h were developed with plot assignment to RHG based on field/office status. There were some instances where RHG were not formed due to small RHG sample sizes ($m_{hg} \leq 5$) that would occur or when the stratum only consisted of either office or field plots. Estimation of forest land area within each estimation unit was accomplished using equations (1)–(7). As estimation units are independent areas, an estimated statewide total forest area and associated variance is obtained by summing over all estimation units. The standard error of the estimate is obtained as the square root of the variance.

For comparative purposes, estimates based on both the simulation data and NFI data from 10 states were produced using current FIA estimation procedures (Scott et al. 2005). Briefly, to account for plots having partial nonresponse, an adjustment factor is calculated for each stratum h and this factor is used to adjust the values for all observed plots (m_{h}) within the stratum.

$$\overline{p}_{\mathrm{h}} = \frac{\sum_{i=1}^{m_{\mathrm{h}}} a_{\mathrm{hi}}}{m_{\mathrm{h}}} \tag{8}$$

$$y'_{\mathrm{hi}} = \frac{y_{\mathrm{hi}}}{\overline{p}_{\mathrm{h}}} \tag{9}$$

RHG are not used and simple random sample estimators are used in each stratum h to obtain a mean and variance of the mean.

$$\overline{y}'_{\mathrm{h}} = \frac{\sum_{i}^{m_{\mathrm{h}}} y'_{\mathrm{hi}}}{m_{\mathrm{h}}} \tag{10}$$

$$v\left(\overline{y}'_{\mathrm{h}}\right) = \frac{\sum_{i}^{m_{\mathrm{h}}} \left(y'_{\mathrm{hi}} - \overline{y}'_{\mathrm{h}}\right)^2}{m_{\mathrm{h}}\left(m_{\mathrm{h}} - 1\right)} \tag{11}$$

Estimates of the population total and corresponding variance are calculated by combining across strata. The formulation of (6) is used to obtain the estimated total ($\hat{y}'$) using $\overline{y}'_h$ values, whereas the variance estimator includes the term for random within-stratum sample sizes that arise due to post-stratification (Scott et al. 2005).

$$v\left(\hat{y}'\right) = \frac{N^2}{m}\left[\sum_{h=1}^{H} W_{\mathrm{h}} m_{\mathrm{h}} v\left(\overline{y}'_{\mathrm{h}}\right) + \sum_{h=1}^{H}\left(1 - W_{\mathrm{h}}\right)\frac{m_{\mathrm{h}}}{m} v\left(\overline{y}'_{\mathrm{h}}\right)\right] \tag{12}$$

Results

The development of the estimation method was primarily motivated to minimize bias in estimates of population totals that arises due to nonresponse. Under the assumption of an equal response probability for each sample plot within an RHG, the mean of the simulation population total estimates was $\hat{y} = 59{,}862{,}329$. Given the true population total of 59,884,824, the bias due to nonresponse was essentially eliminated. To gauge the advantage, the typical implementation of the post-stratified estimator where RHG are not used and it is assumed nonresponse plots are on average equal to their stratum h mean (Scott et al. 2005) produced a mean population total estimate from the simulation of $\hat{y}' = 58{,}148{,}736$, which represents and underestimation of 2.9% compared to the known population mean.

It is also of considerable importance to obtain appropriate estimates of variance to establish valid statistical inference. One concern was the handling of partial nonresponse, which motivated the use of the ratio-to-size estimator at the RHG level (Westfall et al. 2021). The results from $R = 10{,}000$ replications showed a mean estimated variance from (7) of $v(\hat{y}) = 3.67\mathrm{E}12$, which compares favorably with the simulation variance of the population totals being 3.66E12 (0.3% difference). Similarly, the current FIA approach resulted in a mean estimated variance from (12) of $v\left(\hat{y}'\right) = 4.17\mathrm{E}12$ and the variance of the estimated population totals was 3.98E12. This result is consistent with other studies that suggest overestimation of the variance occurs when using (12) in the presence of partial nonresponse in strata comprised primarily of forested land (Westfall et al. 2021; Van Deusen 2005).

Applying the proposed estimation method to FIA data from 10 states resulted in increased forestland area estimates ranging from 0.1 to 3.6% when compared to current FIA estimation methods (Table 2). States having relatively small amounts of forestland (South Dakota, Iowa, and Illinois) exhibited the largest percent increases as creation of RHGs isolated the relatively few field plots and thus nonresponse field plots were likely to better align with the RHG mean than the stratum h mean that would be primarily composed of office plots. Conversely, smaller increases were found in

Table 2 Forest area percent, number of FIA plots, nonresponse percentage on all plots and field plots bases, and forest area change percent and area (ha) for 10 states in the Northern FIA region

State	Forest %	No. plots	Nonresponse %		Forest area	
			All plots	Field plots	Change %	Change (ha)
New York	60.2	5300	12.8	17.1	1.6	1,18,252
West Virginia	78.3	2953	18.3	20.4	1.0	47,585
South Dakota	2.6	8496	0.3	4.1	2.9	22,254
Ohio	30.3	4546	7.5	16.7	2.7	86,492
Iowa	8.2	6067	1.3	10.3	3.0	34,352
Minnesota	33.1	18,051	2.3	6.0	1.1	75,386
Illinois	14.0	6172	3.0	14.0	3.6	71,097
Maine	82.7	3593	2.1	2.4	0.1	5,221
Missouri	34.8	7838	4.5	9.7	1.6	1,01,391
Michigan	54.9	7299	8.7	12.3	1.8	1,48,451
All	34.3	70,315	4.7	9.9	1.5	7,10,482

states where a large portion of the area is forestland (Maine, New York, and West Virginia). However, with the exception of Maine, the potential increases in estimated forest area may be considered substantial as amounts ranging from approximately 20,000 to 150,000 ha were found at the state level and over 700,000 ha in the ten states combined. The standard errors were very similar for both estimation approaches, with those arising from the use of (5) being slightly larger than from (7) due to the estimated RHG weights and accounting for nonresponse uncertainty. As reported in Westfall et al. (2021), bias in variance estimation is small when the adjustment factor ($\overline{p}_h$) is near 1.0. The mean $\overline{p}_h$ across the ten states was 0.988. Although little practical impact of using the ratio-to-size variance estimator (3) within RHG was found in this example, its use is still generally recommended due to the unbiased behavior across a range of partial nonresponse outcomes.

Discussion

The RHG approach relies on the assumption of equal nonresponse probabilities within each RHG. The simulation study implemented this assumption by randomly (equal probability) choosing nonresponse plots at each iteration. Thus, the near equality of the simulation results and the known population total was achieved. In practice, this assumption is easily met for RHG composed of office plots. Meeting this assumption for field plot RHG is likely much more difficult due to the many factors that may contribute to complete nonresponse outcomes. Another important factor that clearly affects nonresponse probabilities is public/private land ownership at the plot location. In the FIA data, estimation units were defined based on public/private ownership maps and thus consideration of this factor for RHGs was not necessary. However, Goeking and Patterson (2013) used more generalized estimation units and inclusion of ownership status was an important part of their RHG construction methods. Still, no assurance is given that nonresponse bias is eliminated by invoking an RHG approach. Minimizing nonresponse bias will depend on how the RHG are assembled and the degree to which the allocation method results in groups of plots with similar nonresponse probabilities. Thus, it is important for the survey practitioner to undertake similar analyses that reflect factors specific to their inventory nonresponse to aid in the development of an effective RHG strategy.

The simulation designed to validate the estimators specified a rate of 10% complete nonresponse and 3% partial nonresponse applied to plots designated as field visit. While the performance of the proposed estimators was acceptable, the amount of bias incurred when using current FIA estimation methods was shown to be 2.9%. Intuitively, larger rates of either complete nonresponse (the primary culprit) or partial nonresponse would result in increasing bias. To validate this supposition, a simulation was also conducted using current methods and a 20% complete nonresponse rate that resulted in the bias increasing to 5.7% underestimation of forest area. Thus, the actual bias that may be present needs to be assessed in the context of the nonresponse rate occurring in the area of interest. The bias also depends on the characteristics of both the population and the strata therein. Further investigation may be warranted to examine how results are affected by various response distribution assumptions and post-stratification scenarios.

It is important to note a key difference in implementation of RHG in this study as compared to the approach taken by Goeking and Patterson (2013). Although in both cases the RHGs were created within already constructed post-strata, Goeking and Patterson (2013) assimilated the RHGs into the original post-stratification such that that total number of strata was increased beyond the original h strata. The weights for the assimilated strata are a product of the known weight for the original post-stratum h and the estimated weight for the assimilated RHG stratum g (step 4 in Goeking and Patterson, 2013) such that the sum of stratum weights within an estimation unit is equal to one. As posited in this study, the original post-stratification was retained and RHGs served as substrata to the existing h strata, i.e., the sum of the weights w_{hg} equals one within each stratum h. Thus, to apply the variance estimators from this study to the strata in Goeking and Patterson (2013), the assimilation process needs to be undone to the original h post-strata, and for the post-strata that were split using RHGs, treat the RHGs as substrata as was done in this study.

Another issue to be considered is development of RHG when samples sizes are small. Westfall et al. (2011) suggest a minimum of 10 plots per stratum to obtain accurate estimates of variance, and Goeking and Patterson (2013) only implemented RHGs if at least 10 plots per RHG could be attained. However, consider a situation where a few field plots (e.g., 5) may exist within a stratum largely composed of office plots (e.g., 95) and one of the field plots is complete nonresponse. Ignoring RHG would result in an estimated stratum mean forest proportion of $4/99 = 0.0404$, whereas implementation of RHG would result in a stratum mean of $0.05 \times 1.0 + 0.95 \times 0.0 = 0.05$. Thus, nearly 25% more forestland would be estimated for the stratum if RHG were implemented. While the actual effect on the population estimate would depend on the stratum weight and possibly other factors, practitioners should weigh the consequences of minimizing nonresponse bias vs obtaining accurate stratum variances when making RHG implementation decisions.

Conclusion

Practically, the amount of nonresponse bias cannot be accurately determined nor is it possible to ensure the bias has largely been removed through various methods thought to address it. However, it is sensible to take steps that mitigate obvious nonresponse bias sources and thus minimize bias in the sample estimate to the extent possible. In this study, differential nonresponse rates due to public/private land ownership were addressed in development of the estimation units. To effectively manage unequal nonresponse rates between field and office plots, RHGs were implemented within existing post-strata. The primary challenge in the RHG approach is estimating the variance, which was accomplished via a double sampling for post-stratification estimator along with an additional term for randomness of nonresponse outcomes.

Comparison of current estimation methods to the RHG approach using NFI data from 10 states showed an increase in estimated forestland area of approximately 1.5% using RHG estimation. This outcome suggests that failure to account for office versus field plot nonresponse differences results in estimates that are too small. The same phenomenon is likely to occur for other attributes such as volume and biomass, although it is acknowledged the direction and magnitude of bias in the estimate depends on factors such as the attribute of interest and the structure of the estimation units and strata therein. As NFI data are often used at various spatial scales for numerous information needs, accurate estimation of forest resources is needed for effective decision making in management and policy arenas. Further, national-scale assessments that gauge contributions to carbon sequestration and other issues of international concern often rely on NFI-based estimates. Estimation processes should be reviewed and modified, if necessary, for forest inventories subject to nonresponse.

Acknowledgements The author is indebted to Dr. Paul L. Patterson for suggesting valuable improvements to the initial draft manuscript. The author is also grateful to the anonymous reviewers for their thoughtful comments to enhance the clarity and technical results presentation.

Funding Not applicable.

Availability of data and material FIA data are publicly accessible; synthetic data on request.

Code availability Upon request.

Declarations

Conflicts of interest None.

References

Cochran WG (1977) Sampling techniques. John Wiley & Sons, New York

Domke GM, Woodall CW, Walters BF, McRoberts RE, Hatfield MA (2014) 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 315:112–120

Goeking SA, Patterson PL (2013) Stratifying to reduce bias caused by high nonresponse rates: A case study from New Mexico's forest inventory. Res. Note RMRS-RN-59. USDA, Forest Service, Rocky Mountain Research Station, Fort Collins, CO

Gormanson DD, Pugh SA, Barnett CJ, Miles PD, Morin RS, Sowers PA, Westfall, JA (2018) Statistics and quality assurance for the Northern Research Station Forest Inventory and Analysis program. Gen. Tech. Rep. NRS-178. USDA, Forest Service, Northern Research Station, Newtown Square, PA

Magnussen S, Stinson G, Boudewyn P (2017) Updating Canada's national forest inventory with multiple imputations of missing contemporary data. For Chron 93(3):213–225

Patterson PL, Coulston JW, Roesch FA, Westfall JA, Hill AD (2012) A primer for nonresponse in the US forest inventory and analysis program. Env Mon Assess 184(3):1423–1433

Särndal C-E, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer-Verlag, New York

Scott CT, Bechtold WA, Reams GA, Smith WD, Westfall JA, Hansen MH, Moisen GG (2005) Sample-based estimators used by the forest inventory and analysis national information management system. Gen. Tech. Rep. SRS-80. USDA, Forest Service, Southern Research Station, Asheville, NC

Van Deusen PC (2005) Comparing mapped plot estimators. In: McRoberts RE, Reams GA, Van Deusen PC, McWilliams WH (eds) Proceedings of the sixth annual forest inventory and analysis symposium. Gen. Tech. Rep. WO-70. USDA, Forest Service, Washington Office, Washington DC

Westfall JA, Patterson PL, Coulston JW (2011) Post-stratified estimation: within-strata and total sample size recommendations. Can J for Res 41(5):1130–1139

Westfall JA, Lister AJ, Scott CT, Weber TA (2019) Double sampling for post-stratification in forest inventory. Eur J for Res 138(3):375–382

Westfall JA, Lister AJ, Scott CT (2021) Evaluation of mapped-plot variance estimators across a range of partial nonresponse in a post-stratified national forest inventory. Can J For Res In press