NOTE / NOTE

Post-stratified estimation: within-strata and total sample size recommendations

James A. Westfall, Paul L. Patterson, and John W. Coulston

Abstract: Post-stratification is used to reduce the variance of estimates of the mean. Because the stratification is not fixed in advance, within-strata sample sizes can be quite small. The survey statistics literature provides some guidance on minimum within-strata sample sizes; however, the recommendations and justifications are inconsistent and apply broadly for many different population structures. The impacts of minimum within-strata and total sample sizes on estimates of means and standard errors were examined for two forest inventory variables: proportion forestland and cubic net volume. Estimates of the means seem unbiased across a range of minimum within-strata sample sizes. A ratio that described the decrease in variability with increasing sample size allowed for assessment of minimum within-strata sample requirements to obtain stable estimates of means. This metric indicated that the minimum within-strata sample size should be at least 10. Estimates of standard errors were found to be biased at small total sample sizes. To obtain a bias of less than 3%, the required minimum total sample size was 25 for proportion forestland and 75 for cubic net volume. The results presented allow analysts to determine within-stratum and total sample size requirements corresponding to their criteria for acceptable levels of bias and variability.

Résumé : La post-stratification est utilisée pour réduire la variance des estimations de la moyenne. Comme la stratification n'est pas fixée à l'avance, la taille de l'échantillon dans les strates peut être très faible. La littérature qui porte sur les résultats d'enquête fournit des indications sur la taille minimale de l'échantillon dans les strates. Mais les recommandations et les justifications sont inconsistantes et s'appliquent de façon générale à plusieurs structures de population différentes. Les impacts de la taille minimale de l'échantillon dans les strates et de la taille totale de l'échantillon sur les estimations de la moyenne et de l'erreur-type ont été examinés pour deux variables d'inventaire forestier : la proportion de terres forestières et le volume net de bois. Les estimations de la moyenne semblent non biaisées pour un éventail de tailles minimales de l'échantillon dans les strates. Une métrique qui décrit la diminution de la variabilité avec l'augmentation de la taille de l'échantillon a permis de déterminer la taille minimale de l'échantillon dans les strates pour obtenir des estimations stables de la moyenne. La métrique a montré que la taille minimale de l'échantillon dans les strates devrait être d'au moins 10. Les estimations de l'erreur-type étaient biaisées lorsque la taille totale de l'échantillon était petite. La taille totale minimale de l'échantillon requise pour obtenir un biais inférieur à 3 % était de 25 pour la proportion de terres forestières et de 75 pour le volume net de bois. Ces résultats permettent aux analystes de déterminer la taille totale de l'échantillon et la taille de l'échantillon dans la strate qui correspondent à leurs critères pour obtenir des niveaux acceptables de biais et variabilité.

[Traduit par la Rédaction]

Introduction

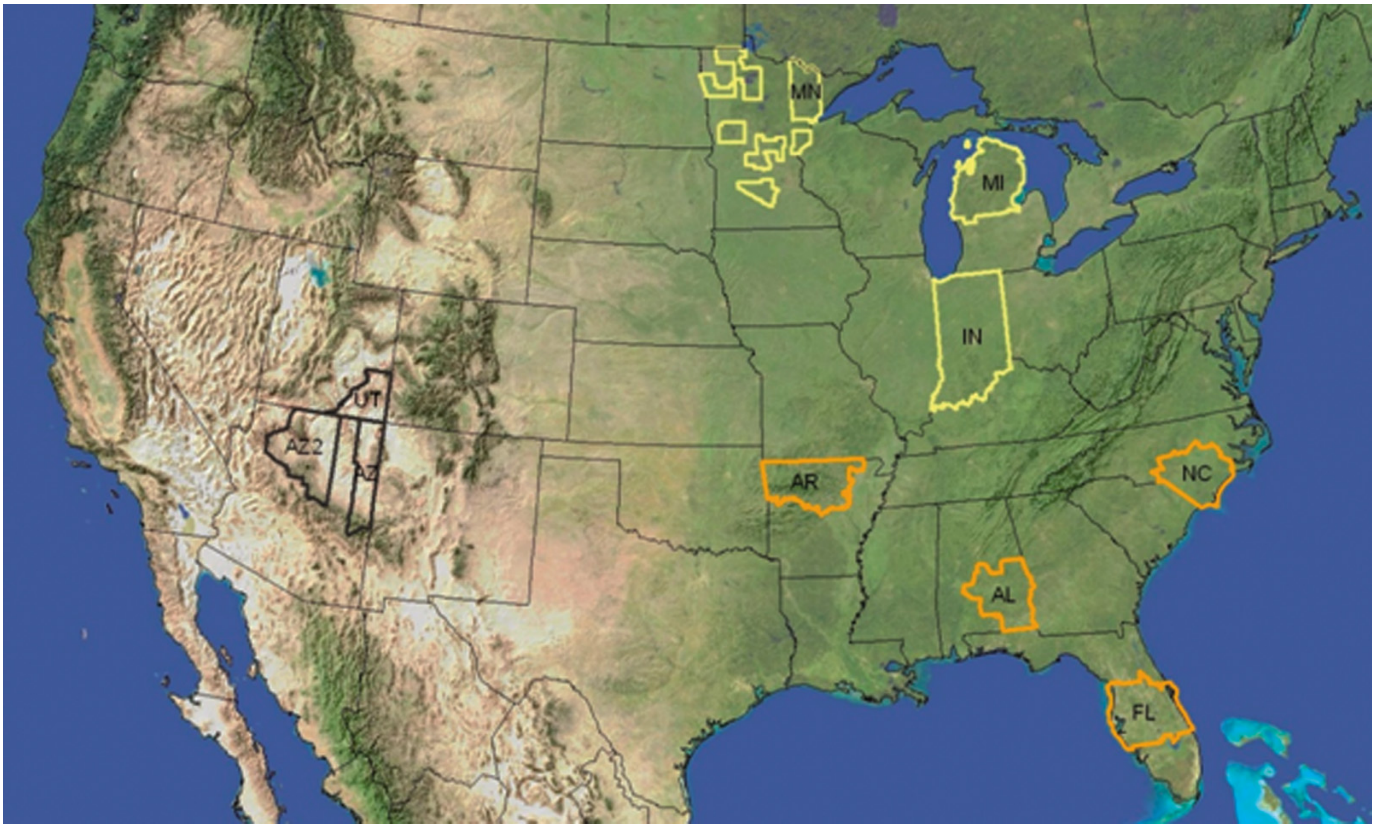
Stratification is a technique used in sample survey design to improve the precision of population estimates (Holt and Smith 1979). In stratified sampling, units of the population having similar characteristics are subdivided into classes (strata). For stratified designs in continuous forest inventory and monitoring programs, this is often accomplished using map-based data (McRoberts and Hansen 1999; Katila and Tomppo 2002). In practice, most strata are developed from wall-to-wall geospatial data (e.g., classified satellite imagery) because it is less time consuming than point-based methods such as double sampling for stratification and there is no additional uncertainty due to estimating the stratum weights (Scott et al. 2005). Continuous forest inventory and monitoring programs usually have fixed plot locations that do not change over time (Brassel and Lischke 2001; Reams et al. 2005). However, the landscape conditions change due to urbanization, forest reversions, etc., and the stratification maps change with the collection and classification of new remotely

Received 4 January 2011. Accepted 10 February 2011. Published at www.nrcresearchpress.com/cjfr on 28 April 2011.

J.A. Westfall. US Forest Service, Northern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073, USA.
P.L. Patterson. US Forest Service, Rocky Mountain Research Station, 2150A Centre Ave., Suite 350, Fort Collins, CO 80526, USA.
J.W. Coulston. US Forest Service, Southern Research Station, 4700 Old Kingston Pike, Knoxville, TN 37919, USA.

Corresponding author: J.A. Westfall (e-mail: jameswestfall@fs.fed.us).

Fig. 1. Map of 10 estimation units in the United States colored by general stratification scheme. The yellow outline (MN, MI, and IN) denotes a canopy cover based stratification, the orange outline (AR, AL, and NC) denotes a fragmentation based stratification, and the heavy black outline (WT, AZ, and AZ2) denotes a forest – nonforest stratification.



sensed data. These changes result in differing stratum boundaries over time. Thus, effective implementation of stratified estimation requires strata delineation after the plot selection has occurred (post-stratification).

While post-stratification is an effective tool for reducing variance, the allocation of sample units to strata cannot be controlled. The variation of sample sizes with respect to strata can result in too few samples to characterize a reliable stratum mean and standard error. Särndal et al. (1992) suggested at least 20 sample units per stratum to obtain stable estimates of stratum means. According to Cochran (1977), post-stratification provides a sample mean that is almost as precise as the sample mean from proportional stratified sampling as long as each stratum has 20 or more sample units, while Kish (1965) recommended a minimum of 10 sample units based on knowledge of the correct stratum weights. These references show disagreement on the minimum number of sample units per stratum and ambiguity in the justification. In addition, these recommendations are intended to apply broadly to all populations, but in practice, the minimum sample size is dependent on the structure of the population. The literature provides no direct guidance on sample size requirements for post-stratified forest populations, particularly for broad-scale forest inventories.

Minimum sample size requirements influence the areal size of a population for which estimates can be made. For example, Cochran (1977) recommended four to six strata with at least 20 sample units per stratum. Based on this recommendation, a minimum of 80–120 sample units would be needed. Because the sampling intensity is fixed in most large-area forest inventories, these criteria constrain populations to certain minimum area requirements. For example, the Forest Inventory and Analysis (FIA) program of the US Forest Service typically defines populations by county boundaries and the areal extent of these populations range from less than 20 000 ha to more than 5 000 000 ha (US Census Bureau 2000). To implement the suggested criteria, given the FIA sampling intensity of approximately 1 plot per 2430 ha, the minimum areal extent of a population is approximately 194 400 – 291 600 ha. In the coterminous United States, these thresholds are met by only 37% and 17% of the counties, respectively. The remaining 63%–83% of the counties would be combined in some fashion to meet the recommended sample size requirements. This is problematic because combining counties into larger populations results in disagreement between the area of each county reported by the US Census Bureau (2000) and the estimated area of each county from the FIA program. This provides the motivation for examining minimum sample size requirements in more detail than provided by the survey statistics literature.

In this paper, we examine minimum sample size requirements for forest populations in the United States. To assess the sample size requirements, repeated samples from populations of plots were drawn for varying sample sizes. For each sample size, the properties of the distributions were analyzed to assess where stability of the estimates occurred. The attributes of interest were forestland area and net cubic volume, as these are often the primary measures for current status of forest resources.

Data

The populations used for analyses were constructed from ground plot data measured by FIA under the annual inventory system (McRoberts 2005). The plots are spatially distributed via a hexagonal grid with a distance of approximately 5 km between hex centers. A quasi-systematic sample results from one plot being randomly located within each hexagon. The mapped plot design used by FIA provides information on within-plot boundaries where conditions differ among various attributes (US Forest Service 2007). Knowledge of these boundaries allows for computation of the proportion of forestland on each plot. Cubic volumes for merchantable trees (diameter at breast height 12.7 cm and larger) were computed from regionally specific, individual-tree models that utilize various tree size information, e.g., diameter at breast height. Plot volumes were calculated by summing individual-tree volumes.

Ten land areas of interest were identified for analysis (Fig. 1). These areas were in the states of Alabama (AL), Arizona (AZ and AZ2), Arkansas (AR), Florida (FL), Indiana (IN), Missouri (MO), Minnesota (MN), North Carolina (NC), and Utah (UT) in the United States. Each of these areas constitutes a region in which estimates of population parameters are desired and, for the purposes of this paper, will be referred to as estimation units. These estimation units were chosen to represent a wide range of environmental conditions.

Each estimation unit was post-stratified using one of three general schemes based on a wall-to-wall map. In each case, plots were assigned to a map-based class via spatial overlay of plot coordinates. The three general classification schemes were canopy cover based on the National Land Cover Database canopy cover map (Homer et al. 2004), fragmentation based on the National Land Cover Database land cover map (Homer et al. 2004), and forest – nonforest based on the map develop by Ruefenacht et al. (2008). For each estimation unit, Table 1 indicates the specific stratification scheme that was applied to each estimation unit as well as providing information regarding pertinent characteristics.

Methods

For proportion of forested area and cubic net volume attributes, we developed a resampling simulation study to evaluate the behavior of estimates for the mean and standard error under various sample size scenarios. For each estimation unit, the plots were considered to be the population from which samples were drawn (Arner et al. 2004) and the stratum membership of each plot was known. To minimize complexity, the simulations were initially conducted using proportional allocation of plots to strata. To mimic exact proportional allocation, the stratum weights within each estimation unit (W_{jh}) were determined by dividing the number of plots in the estimation unit (N_j) by the number of plots in the stratum (N_{jh}), where j is the index of the estimation unit and h is the index of the stratum.

To mimic a forest inventory sample, a random sample (without replacement) was taken from the population of plots within each stratum. The minimum stratum samples sizes considered ranged from 2 to 20 plots and the minimum number (m) was chosen for the stratum having the smallest weight (W_{j0}). For a given minimum strata sample size, the number of plots to sample in the remaining strata was proportional to strata areas:

$$[1] \qquad n_{jh}(m) = \text{round}\left(m\frac{W_{jh}}{W_{j0}}\right)$$

where $n_{jh}(m)$ is the number of plots sampled in estimation unit j, stratum h for minimum stratum sample size m and W_{jh} is the weight for estimation unit j, stratum h. For estimation unit j and minimum sample size m, the total sample size is

$$[2] \qquad n_j(m) = \sum_{h}^{H_j} n_{jh}(m)$$

Using the sample data and the stratum weight information, estimates of the population mean and its standard error were calculated. Although other approaches to stratified estimation with mapped plots in forest inventory have been proposed (e.g., Van Deusen 2005), this study followed the FIA protocols outlined by Scott et al. (2005). Although the stratum assignment of each plot was known in advance of drawing the sample in this study, the data were treated as a post-stratified sample, i.e., the variance estimator includes the additional uncertainty associated with random sample sizes within strata. The estimators for the mean and associated standard error were as follows:

$$[3] \qquad \bar{y}_{jh} = \frac{\sum_{i=1}^{n_{jh}} y_{jhi}}{n_{jh}}$$

$$[4] \qquad \bar{y}_j = \sum_{h=1}^{H} W_{jh}\bar{y}_{jh}$$

$$[5] \qquad v(\bar{y}_j) = \frac{1}{n_j}\left[\sum_{h=1}^{H} W_{jh}S_{jh}^2 + \sum_{h=1}^{H}(1 - W_{jh})\frac{1}{n_j}S_{jh}^2\right]$$

$$[6] \qquad \text{se}(\bar{y}_j) = \sqrt{v(\bar{y}_j)}$$

where $\bar{y}_j$ is the sample mean for estimation unit j, $\text{se}(\bar{y}_j)$ is the estimated standard error of sample mean for estimation unit j, $\bar{y}_{jh}$ is the sample mean of estimation unit j, stratum h, y_{jhi} is the sample unit observation i of estimation unit j, stratum h, and S_{jh}^2 is the estimated variance in estimation j, stratum h where

$$S_{jh}^2 = \frac{\sum_{i=1}^{n_{jh}} (y_{jhi} - \bar{y}_{jh})^2}{(n_{jh} - 1)}$$

and other terms are as previously defined.

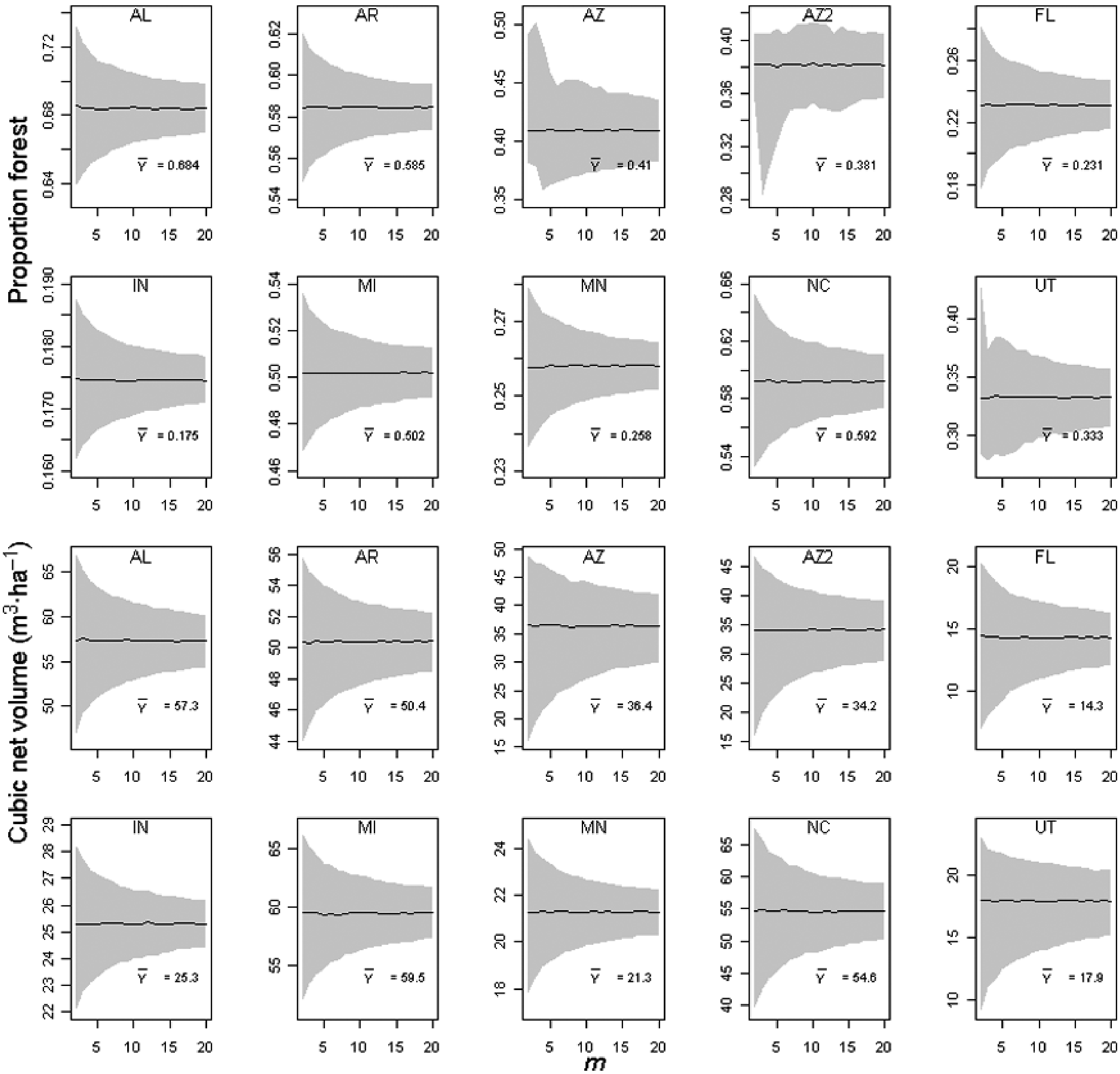
For each estimation unit j, this process was repeated 10 000 times for m = 2 to 20, i.e., the sample mean and estimated standard error were calculated for each sample. These data were used to compute summary statistics that were used

Table 1. Summary of population means (with coefficient of variation percent in parentheses) and specific stratification information for the 10 estimation units in the United States.

Study area	N		Stratum 1	2	3	4	5	Mean
AL[a]	1487	W_h	0.125	0.179	0.238	0.458		
		Proportion forest	0.090 (307)	0.291 (139)	0.746 (48)	0.967 (15)		0.684
		Volume ($m^3 \cdot ha^{-1}$)	1.2 (626)	12.4 (239)	54.4 (105)	91.7 (82)		57.3
NC[a]	1174	W_h	0.219	0.240	0.237	0.304		
		Proportion forest	0.384 (103)	0.819 (35)	0.082 (326)	0.961 (15)		0.592
		Volume ($m^3 \cdot ha^{-1}$)	24.6 (194)	73.3 (98)	2.0 (668)	102.5 (84)		54.6
AR[b]	1756	W_h	0.075	0.103	0.313	0.510		
		Proportion forest	0.363 (114)	0.625 (64)	0.106 (239)	0.903 (27)		0.585
		Volume ($m^3 \cdot ha^{-1}$)	19.8 (170)	34.8 (123)	5.3 (329)	85.7 (75)		50.4
AZ[c]	941	W_h	0.383	0.617				
		Proportion forest	0.882 (34)	0.117 (261)				0.410
		Volume ($m^3 \cdot ha^{-1}$)	89.8 (117)	3.3 (548)				36.4
AZ2[c]	1563	W_h	0.404	0.596				
		Proportion forest	0.860 (38)	0.056 (399)				0.381
		Volume ($m^3 \cdot ha^{-1}$)	82.7 (101)	1.3 (519)				34.2
UT[c]	749	W_h	0.284	0.716				
		Proportion forest	0.857 (39)	0.124 (252)				0.333
		Volume ($m^3 \cdot ha^{-1}$)	53.7 (115)	3.6 (330)				17.9
FL[d]	1955	W_h	0.131	0.170	0.315	0.385		
		Proportion forest	0.001 (1600)	0.761 (52)	0.021 (614)	0.247 (156)		0.231
		Volume ($m^3 \cdot ha^{-1}$)	0.0	58.3 (145)	0.2 (1163)	11.4 (314)		14.3
IN[e]	3704	W_h	0.029	0.041	0.072	0.086	0.773	
		Proportion forest	0.550 (70)	0.233 (138)	0.862 (30)	0.847 (30)	0.019 (512)	0.175
		Volume ($m^3 \cdot ha^{-1}$)	70.2 (107)	34.8 (166)	124.7 (61)	126.0 (62)	2.7 (590)	25.3
MI[e]	2993	W_h	0.076	0.112	0.160	0.219	0.434	
		Proportion forest	0.789 (42)	0.524 (82)	0.905 (25)	0.929 (20)	0.083 (292)	0.502
		Volume ($m^3 \cdot ha^{-1}$)	74.5 (91)	41.0 (122)	111.7 (64)	135.2 (63)	4.3 (383)	59.5
MN[e]	3239	W_h	0.035	0.036	0.118	0.147	0.663	
		Proportion forest	0.202 (181)	0.414 (101)	0.709 (55)	0.857 (34)	0.039 (426)	0.258
		Volume ($m^3 \cdot ha^{-1}$)	8.3 (251)	29.5 (150)	56.1 (115)	78.9 (86)	2.5 (605)	21.3

[a]Fragmentation stratification with forest, forest edge, nonforest, and nonforested edge categories.
[b]Canopy cover stratification with four cover categories (0%–25%, 26%–50%, 51%–75%, and 76%–100%).
[c]Forest – nonforest stratification.
[d]Fragmentation stratification with forest, nonforest, edge, and water categories.
[e]Canopy cover stratification with five cover categories (0%–5%, 6%–50%, 51%–65%, 66%–80%, and 81%–100%).

Fig. 2. Mean (line) and interquartile range (shaded) of sample means for (*a*) proportion forest and (*b*) cubic net volume from 10 000 simulations for $m = 2$ to 20 for 10 estimation units in the United States. The population mean is given by $\overline{Y}$.



to evaluate the changes in the conditional distributions of the means and standard errors as m increases. The distributions of $\bar{y}_j$ and $\mathrm{se}(\bar{y}_j)$ at each m were characterized by the mean and, for $\bar{y}_j$, the interquartile range (IQR). The nonparametric IQR was chosen for the measure of dispersion, as the unknown forms of the underlying distributions of $\bar{y}_j$ precluded adoption of an appropriate parametric estimator. The two primary aspects of interest were (*i*) whether bias in $\mathrm{se}(\bar{y}_j)$ was encountered at small sample sizes and (*ii*) at what point does increasing sample size provide relatively small reductions in variability for $\bar{y}_j$.

For comparisons with simulation results, the true population mean and standard error were computed from all plots within each estimation unit. The population mean ($\overline{Y}_j$) is calculated using all elements of the population; however, to make comparisons of standard errors for the post-stratification estimator, the sample size needs to be taken into account. Thus, the standard error of $\overline{Y}_j$ needs to be computed at each m. The following formula (ignoring the negligible finite population factor) shows the calculation of the standard error from a population perspective, i.e., using the population-level stratum weights and variances (Cochran 1977):

$$[7] \qquad V(\bar{y}_j) = \frac{1}{n_j}\sum_{h}^{H} W_{jh}S_{jh}^2 + \frac{1}{n_j^2}\sum_{h}^{H}(1 - W_{jh})S_{jh}^2$$

where

$$S_{jh}^2 = \frac{\sum_{i}^{N_{jh}}(y_{jhi} - \overline{Y}_{jh})^2}{(N_{jh} - 1)}$$

$$[8] \qquad \mathrm{SE}(\bar{y}_j) = \sqrt{V(\bar{y}_j)}$$

The reduction in the variability of the mean (measured by IQR) as m increases was evaluated via the behavior of

$$[9] \qquad R(m) = \frac{|f'(m)|}{f(m)}100$$

where $f(m)$ is a nonlinear curve fitted over all estimation units that describes the relationship between IQR and m and $f'(m)$ is the slope of line tangent to $f(m)$. If the IQR is decreasing, $R(m)$ measures the rate of decrease as a percentage of the IQR. As $R(m)$ decreases, there are smaller percentage reductions in the IQR for each increase in m. At some point, the percentage reductions become sufficiently small such that increasing m produces inconsequential decreases in IQR.

Since the standard error of the post-stratification estimator decreases as sample size increases, the percent difference between the estimated standard error and the true population standard error was used to investigate any bias of the estimated standard error conditional on the minimum sample size, where the percent difference is

$$[10] \qquad \%D(\mathrm{se})_j = \frac{\mathrm{se}(\bar{y}_j) - \mathrm{SE}(\bar{y}_j)}{\mathrm{SE}(\bar{y}_j)}100$$

The mean of the conditional empirical distribution of $\%D(\mathrm{se})_j$, given the minimum stratum sample size, was used as a measure of the bias of estimated standard error. The population standard error and the estimated standard error and hence the mean of the conditional empirical distribution of $\%D(\mathrm{se})_j$ are functions of many factors, including the minimum stratum sample size, total sample size, number of strata, and homogeneity within the strata. For each minimum stratum sample size m, the relationship between the mean of the conditional empirical distribution of $\%D(\mathrm{se})_j$ and sample size $n_j(m)$ was investigated by fitting a curve over all estimation units. The curve fitting was done using the statistical package R (R Development Core Team 2008).

Additional analyses were conducted using nonproportional allocation to determine if nonproportionality affected the sample size requirements. In earlier analyses, exact proportional allocation was used to avoid introducing any effects due to inaccurate stratum weights. However, this condition rarely exists in practice. To mimic the nonproportional allocation of plots to strata that often occurs due to the combination of post-stratification and sample unit nonresponse (e.g., lack of access or hazardous conditions), the stratum weights were altered. The degree to which the map-based stratum weights differed from proportional allocation was determined from relationships developed through analysis of FIA post-stratification outcomes. The weights corresponding to proportional allocation were computed from the number of sample plots in the stratum divided by the total number of plots in the estimation unit. Specifically, for each map-based stratum weight (rounded to the nearest 0.01), we found the largest (in magnitude) positive and negative percent difference from the proportional weight. The percent differences tended to be large when stratum weights were small and vice versa. There was an observable relationship between the proportional weights and the percent differences, which were described by regression models to predict the maximum positive and negative deviation from the proportional weight likely to be observed with post-stratification. The predicted percent differences were implemented as the upper and lower bounds within which the altered weights must reside. For each estimation unit, weights of randomly selected strata were set at either the upper or the lower bound and remaining strata were adjusted accordingly (within bounds) such that the sum of the weights remained equal to 1.

Results and discussion

The fundamental purpose of forest inventories is to obtain unbiased estimates of means (and their standard errors) for attributes of interest. The results of the simulation study were used to evaluate the properties of the sample mean and estimated standard error over a range of minimum stratum sample sizes (m). The post-stratified estimator and the approximate variance (and (or) its estimator) used in this study are found in most comprehensive sampling literature (e.g., Cochran 1977; Särndal et al. 1992; Gregoire and Valentine 2008). The estimator and estimated variance are close to unbiased when considered over all possible samples with at least one element in each stratum. In this study, we have a set of samples restricted by minimum stratum sample size m and the remainder of the strata sample sizes distributed proportionally per eq. 1. Also, there is the additional restriction that the minimum stratum size occurs in the smallest strata. Thus, the statistical properties of the estimator and estimated variance over this restricted set of samples may be different from those found over the complete set of samples of size $n_j(m)$

To assess possible bias in the sample means, the average of the sample means for m were plotted against m for each estimation unit. Figure 2 depicts the results for estimates of both proportion forestland and cubic volume where it can be seen that no trend is evident for either attribute in any of the estimation units. Also, these averages clustered about the population mean. Thus, it can be concluded that no discernable bias exists in estimates of means for forest area and volume, regardless of stratum sample sizes.

For forestland area and cubic volume, the patterns in IQR were very similar among the estimation units (Fig. 2). However, the two estimation units in AZ and the estimation unit in UT exhibited notably different relationships for forestland area. For these estimation units, the pattern was much more erratic at the smaller m. This result was due to small sample sizes n_j that produce some estimates $\bar{y}_j$ that differ substantially from the population mean ($\overline{Y}_j$). For example, in the estimation unit AZ at m = 2, there would be two units in the smallest stratum and three units in the other stratum, resulting in a total sample size of five units. Regardless of this errati-

Table 2. Values of $R(m)$ for forestland area and cubic volume by m.

	$R(m)$	
m	Forestland area	Cubic volume
2	23.158	24.546
3	15.439	16.364
4	11.579	12.273
5	9.263	9.818
6	7.719	8.182
7	6.617	7.013
8	5.789	6.137
9	5.146	5.455
10	4.632	4.909
11	4.211	4.463
12	3.86	4.091
13	3.563	3.776
14	3.308	3.507
15	3.088	3.273
16	2.895	3.068
17	2.724	2.888
18	2.573	2.727
19	2.438	2.584
20	2.316	2.455

cism, the patterns were similar to the other estimation units in that rates of reduction in IQR decrease with increasing m.

To evaluate where diminishing gains in reduction of variability of the sample means may occur, the value of $R(m)$ was evaluated against m (Table 2). It is shown that the values of $R(m)$ for a given m are quite similar for forestland area and cubic volume attributes. Also, the values of $R(m)$ decrease at a diminishing rate with increasing m. If an acceptable percentage rate of decrease, for instance, was thought to be $R(m) \leq 10.0$, then the minimum within-stratum sample size would be $m = 5$. We considered $R(m) \leq 5.0$ to be a practical (e.g., cost versus precision) threshold for forest inventory purposes (i.e., the rate of decrease in variability is relatively small ($\leq$5%) compared with the magnitude of the IQR). If this threshold is adopted as an indicator of where increasing m provides only marginal decreases in IQR, then a minimum sample size of 10 plots per stratum is warranted (Table 2).

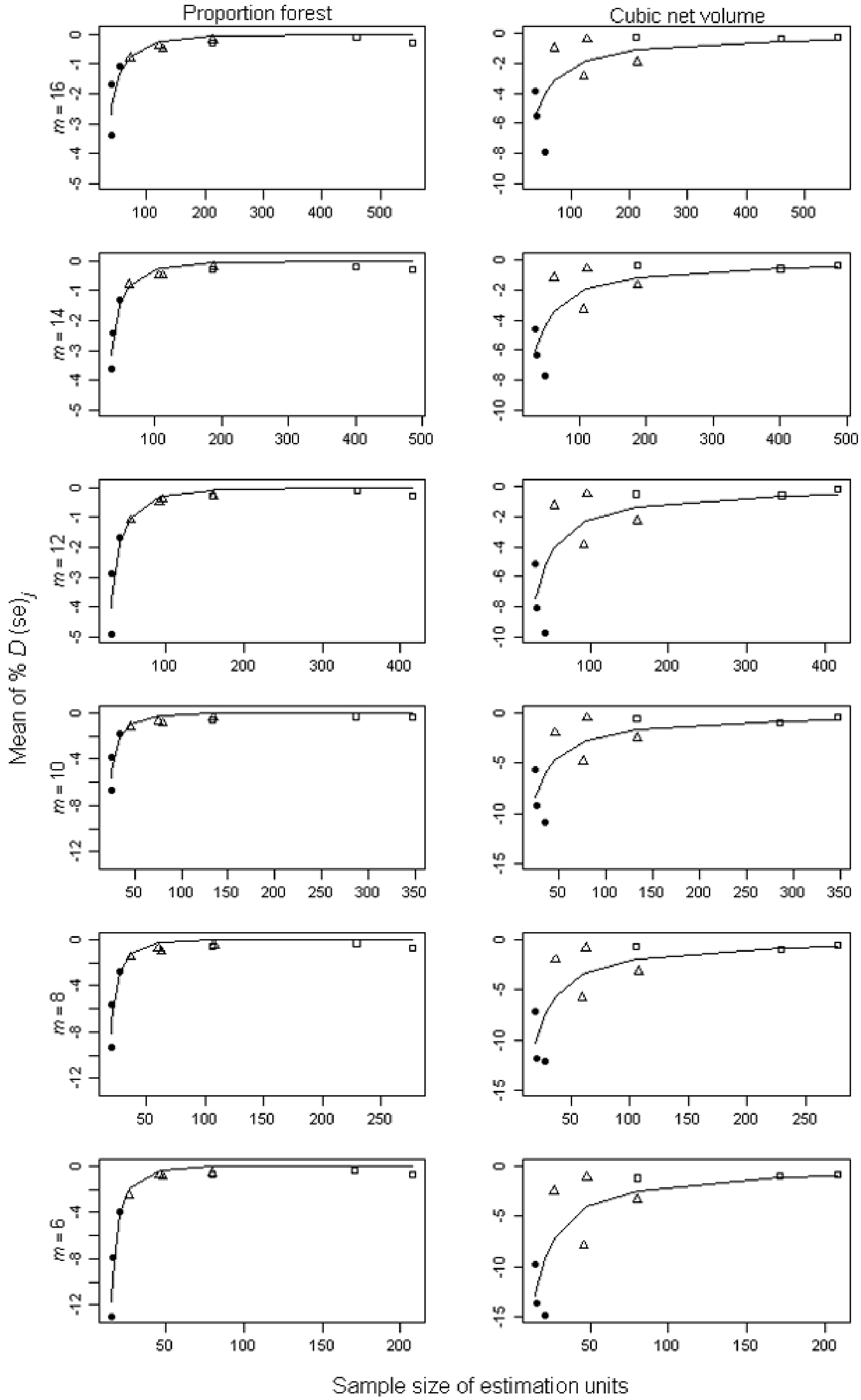
As a measure of bias in the estimated standard error of the mean, the mean of $\%D(\mathrm{se})_j$ was evaluated against the total sample size of the estimation unit, $n_j(m)$ (Fig. 3). To assist in evaluating the effect of total sample size on the bias of the estimated standard error for each minimum sample size m, a curve was fit to the data points. The estimation units represent a range of the number of strata, size of stratum weights, within-strata variability, and total sample size. Figure 3 indicates that for both proportion of forest and cubic net volume, the estimated standard error is too small and for a fixed minimum stratum sample size m, the bias decreases as the sample size increases. Similarly, as m (and n_j) increases, the amount of bias decreases. For a given m, the bias in the estimated standard error appears to be smaller and less variable as a function of total sample size for the proportion of forest than for cubic volume (owing to more strata homogeneity within estimation units). Using Fig. 3, the user can determine the minimum sample size needed to limit the bias of the estimated standard error. For example, for estimates of cubic volume and a minimum stratum sample size of 10, a minimum sample size of 75 units is needed to expect a bias of 3% or less in the estimated standard error. The minimum sample size for proportion forest under the same criteria is approximately 25 units, indicating that required sample sizes are attribute dependent.

For small sample sizes, the negative bias in the standard error results from estimates of stratum variances that are too small. Specifically, this occurs when a stratum is composed primarily of a number of equal (or nearly equal) values, but also some values that deviate considerably from the similar observations. Thus, the stratum is quite homogeneous (which is strived for), with the exception of a few outliers. When taking small samples ($m < 10$) from such strata, the sample will often only contain values that are the same (or similar) and thus the S_{jh}^2 is underestimated. In the context of this study, this phenomenom mostly occurs in the stratum where most nonforest plots reside, i.e., the proportion of forestland area and cubic volume are zero for most plots. However, other scenarios may be envisioned where this phenomenon could occur in other strata. For strata having this type of underlying population structure, the lack of homogeneity results not only in less precise estimates of $\bar{y}_j$ but also in bias in estimated standard errors for sample sizes that are small ($m < 10$).

When strata have a broader distribution of values (or when all values are identical), the underestimation of S_{jh}^2 largely disappears. The issue is also ameliorated if the size of the stratum having a number of similar values and a few dissimilar values is relatively large, as the n_{jh} is likely to exceed 10 even for smaller values of m (recall that m is associated with the smallest stratum and other stratum sample sizes are scaled accordingly). For example, this may occur in populations that have a substantial nonforest land area.

The second set of analyses performed to evaluate whether the results were affected by nonproportional allocation showed that the minimum sample size recommendations given above were still valid. The metrics evaluated were the same as those shown in Fig. 2 and there were no trends in the estimate of $\bar{y}_j$ and the rate of reduction in variability (as measured by IQR) was comparable with the earlier analyses. Similar standard errors were also noted. The estimates of $\bar{y}_j$ were consistent with the given stratum weights; however, these estimates differed, sometimes substantially, from $\overline{Y}_j$. This is because the nonproportionality brings the effect of stratification into the estimation (Kish 1965). Whether this adjustment provides more accurate estimation of the population mean depends on the actual values of the nonsampled units. This phenomenon should be of concern, as the estimated means may be incongruous with those obtained from a proportional sample (which would be equivalent to a simple random sample.). For example, the mean volume from a forest inventory is estimated under a simple random sampling design. Subsequently, post-stratification is implemented in an attempt to decrease the standard error, but the sample units do not distibute proportionately with the stratum weights. While both the stratified and the simple random sample estimators are theoretically unbiased, the two estimates of mean volume may differ by a considerable amount. As such, it may be advisable to implement measures that ensure that the sam-

Fig. 3. Plots of estimation unit mean %D(se) versus estimation unit sample sizes for minimum stratum sample sizes m = 6, 8, 10, 12, 14, and 16. The line depicts the average trend estimated from a nonlinear regression model. Markers denote the number of strata in the estimation unit (circles, two strata; triangles, four strata; squares, five strata).



ple units are nearly proportionally distributed under low levels of nonresponse (Coulston 2008).

Most large-area forest inventories are conducted using a systematic sample design. In this study, simulation populations were constructed using the quasi-systematic FIA data and sample units were randomly selected from these populations. There may be some samples where the plots are geographically clustered, although on average are at least 5 km apart (but usually much farther). Since the simulation population is from a quasi-systematic sample, the standard error based on the simple random samples is slightly higher than, but approximately equivalent to, the standard error based on systematic samples (based on work not shown here). This also applies to our results, which pertain to subsets of the sample space, i.e., samples with fixed stratum sample sizes. Thus, the recommendations are also applicable to systematic samples.

Although these results were obtained using certain populations in the United States, it is expected that similar outcomes would be realized from forest inventories of other populations as well. However, there may be situations where the populations are of sufficiently different structure and these results may not apply. For example, a population defined strictly as forestland area at the time of the inventory would probably have different characteristics, e.g., it would be unlikely to have a stratum where a large proportion of observed data points for cubic volume would be zeros. It should also be noted that although we used a varying number of strata within our populations and the estimation unit sample sizes n_j were dependent on the number strata and the distribution of stratum weights, we are not specifically making recommendations on the number of strata that should be used.

Conclusions

Post-stratification is an important technique for variance reduction in forest inventories that maintain fixed sample plot locations for multiple measurement cycles. However, implementation must be carefully considered to obtain precise and accurate statistics, particularly in the context of both the within-stratum and total sample sizes. In this paper, we found that estimates for the population mean of both proportion forestland and cubic volume appear to be unbiased regardless of sample size. There was substantial variability in the means for small sample sizes, although this variability decreased relatively quickly as sample size increased. Using the threshold $R(m) \leq 5.0$ as an indicator of where precision gains begin to fade with increasing sample size suggests that 10 sample units in a stratum is sufficient for obtaining a stable estimate of the mean.

For small sample sizes, estimates of the standard error were found to be too small when compared with the true population value. It was discovered that the bias was related to both the minimum strata sample size and the total sample size for the estimation unit, requiring both criteria to be taken into account for estimation in a post-stratified design. When the number of strata is small, a minimum within-strata sample size criterion is often obtainable, while the total sample size may still be quite small such that considerable bias in the estimated standard error is still present. Conversely, when a relatively large number of strata are present, the total sample size tends to be large, but small within-stratum sample sizes can result in appreciable standard error bias. Furthermore, there were differences between proportion of forestland and cubic volume attributes, with cubic volume requiring a larger sample size to obtain negligible standard error bias. For inventories utilizing a single post-stratification for both attributes, the sample sizes associated with cubic volume are needed to ensure appropriate estimates of standard errors for both attributes.

These results provide guidance for those using a post-stratified forest inventory design. We recommend that within-strata sample sizes of 10 observations are sufficient for stability of the mean and the estimated standard error of the mean when post-stratification is employed. At this minimum stratum sample size, we also suggest total sample sizes of at least 25 units for proportion forestland and approximately 75 units for cubic volume to obtain estimates of the standard error with a maximum of 3% bias. Under such criteria, the strategy would be to construct estimation units of sufficient areal extent to capture the minimum total sample size and then devise a post-stratification scheme in which the smallest within-stratum sample size is at least 10 units. We recognize that other analysts may have differing opinions of acceptable bias and the results provided herein can be used to determine sample size requirements for their respective viewpoints.

Acknowledgements

This research was funded by the US Forest Service Forest Inventory and Analysis program. The authors are grateful to Dr. Kim Iles for comments that substantially improved the manuscript.

References

Arner, S.L., Westfall, J.A., and Scott, C.T. 2004. Comparison of annual inventory designs using forest inventory and analysis data. For. Sci. **50**: 188–203.

Brassel, P., and Lischke, H. (*Editors*). 2001. Swiss national forest inventory: methods and models of the second assessment. WSL Swiss Federal Research Institute, Birmensdorf.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. John Wiley & Sons, New York.

Coulston, J. 2008. Forest inventory and stratified estimation: a cautionary note. U.S. For. Serv. Res. Note SRS-16.

Gregoire, T.G., and Valentine, H.T. 2008. Sampling strategies for natural resources and the environment. Chapman & Hall/CRC, New York.

Holt, D., and Smith, T.M.F. 1979. Post stratification. J. R. Stat. Soc. Ser. A, **142**: 33–46.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M. 2004. Development of a 2001 national landcover vatabase for the United States. Photogramm. Eng. Remote Sens. **70**: 829–840.

Katila, M., and Tomppo, E. 2002. Stratification by ancillary data in multisource forest inventories employing *k*-nearest neighbour estimation. Can. J. For. Res. **32**(9): 1548–1561. doi:10.1139/X02-047.

Kish, L. 1965. Survey sampling. John Wiley & Sons, New York.

McRoberts, R.E. 2005. The enhanced Forest Inventory and Analysis program. *In* The enhanced Forest Inventory and Analysis program — national sampling design and estimation procedures. *Edited* by W.A. Bechtold and P.L. Patterson. U.S. For. Serv. Gen. Tech. Rep. SRS-80. pp. 1–10.

McRoberts, R.E., and Hansen, M.H. 1999. Annual forest inventories for the north central region of the United States. J. Agric. Biol. Environ. Stat. **4**(4): 361–371. doi:10.2307/1400495.

R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reams, G.A., Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A., and Moisen, G.G. 2005. The Forest Inventory and Analysis sampling frame. *In* The enhanced Forest Inventory and Analysis program — national sampling design and estimation procedures. *Edited* by W.A. Bechtold and P.L. Patterson. U.S. For. Serv. Gen. Tech. Rep. SRS-80. pp. 11–26.

Ruefenacht, B., Finco, M.V., Nelson, M.D., Czaplewski, R., Helmer, E.H., Blackard, J.A., Holden, G.R., Lister, A.J., Salajanu, D., Weyermann, D., and Winterberger, K. 2008. Conterminous US and Alaska forest type mapping using forest inventory and analysis data. Photogramm. Eng. Remote Sens. **74**: 1379–1388.

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

Scott, C.T., Bechtold, W.A., Reams, G.A., Smith, W.D., Westfall, J. A., Hansen, M.H., and Moisen, G.G. 2005. Sample-based estimators used by the Forest Inventory and Analysis national information management system. *In* The enhanced Forest Inventory and Analysis program — national sampling design and estimation procedures. *Edited* by W.A. Bechtold and P.L. Patterson. U.S. For. Serv. Gen. Tech. Rep. SRS-80. pp. 43–67.

US Census Bureau. 2000. Census 2000 TIGER/Line Files [machine-readable data files]. U.S. Department of Commerce, Census Bureau, Washington, D.C.

US Forest Service. 2007. Forest inventory and analysis national core field guide. Vol. 1. Field data collection procedures for phase 2 plots, version 4.0 [online]. U.S. Department of Agriculture, Forest Service, Washington, D.C. Available from http://fia.fs.fed.us/library/field-guides-methods-proc/docs/core_ver_4-0_10_2007_p2.pdf [accessed 04/23/2010].

Van Deusen, P.C. 2005. Stratified forest inventory estimation with mapped plots. Can. J. For. Res. **35**(10): 2382–2386. doi:10.1139/X05-147.