



Sample strategies for bias correction of regional LiDAR-assisted forest inventory Estimates on small woodlots

Yung-Han Hsu¹ · Yingbing Chen¹ · Ting-Ru Yang¹ · John A. Kershaw Jr.¹ · Mark J. Ducey²

Received: 15 December 2019 / Accepted: 16 June 2020 / Published online: 28 July 2020
© INRAE and Springer-Verlag France SAS, part of Springer Nature 2020

Abstract

• **Key message** This study presents an easy-to-apply variable probability sample design that is an efficient and cost-effective method to correct for local bias in regional LiDAR-assisted forest inventory estimates. This design is especially useful for small woodlot owners.

• **Context** Light detection and ranging (LiDAR)-derived forest inventory estimates are generally unbiased at landscape levels but may be biased locally. One solution to correct local bias is to use ground-based double sampling with ratio estimation where the LiDAR estimates form the large sample covariate and the ground plots are used to estimate a correction or calibration ratio.

• **Aims** Our objectives were to test the performance of different sample strategies, to correct for local bias, and to determine the most efficient and cost-effective sampling design.

• **Methods** We compared five sample selection methods and four plot types using simulation. Sample sizes and inventory costs required to achieve 5% standard error were calculated to assess sampling efficiency.

• **Results** The results showed that bias can be corrected successfully using a doubling sampling approach with ratio estimation, and that variable probability selection methods were more efficient than equal probability selection methods. A big basal area factor (BAF) plot was the most cost-effective on-the-ground plot type.

• **Conclusion** The most efficient and cost-effective sampling design was list sampling with big BAF plots. This combination can be used to calibrate LiDAR-derived forest inventory estimates for a variety of forest attributes.

Keywords LiDAR-assisted inventory · Variable probability sampling · Big BAF sampling · Ratio estimation · Sampling with covariates · Sampling to correct

Handling Editor: Jean-Michel Leban

Authors' contributions

Conceptualization: John A. Kershaw;
Formal Analysis: Yung-Han Hsu;
Writing—original draft: Yung-Han Hsu;
Writing—review and editing: John A. Kershaw, Ting-Ru Yang, Yingbing Chen, Mark Ducey;
Supervision: John Kershaw

✉ Yung-Han Hsu
yhsu@unb.ca

Yingbing Chen
ychen22@unb.ca

Ting-Ru Yang
b00605080@gmail.com

John A. Kershaw, Jr.
kershaw@unb.ca

Mark J. Ducey
mark.ducey@unh.edu

¹ Faculty of Forestry and Environment Management, University of New Brunswick, PO Box 4400, Fredericton, NB E3B 5A3, Canada

² Department of Natural Resources and the Environment, University of New Hampshire, 114 James Hall, 56 College Road, Durham, NH 03824, USA

1 Introduction

Airborne light detection and ranging (LiDAR) scanning (ALS) is a remote sensing technique that can provide three-dimensional spatial information about forest canopies and the underlying terrain. Based on this information, ALS is used to produce estimates of forest attributes, such as tree height, diameter, crown width, and volume (Næsset 2002; Corona and Fattorini 2008; Asner et al. 2011; Hayashi et al. 2015). These estimates are commonly referred to as Enhanced Forest Inventory (EFI) in Canada (Canada 2013).

In New Brunswick, the provincial government is in the process of producing freely available EFI estimates for the entire province (Dick 2019). The advantage of EFI is that it provides high-resolution spatial forest inventory estimates across the landscape (Andersen et al. 2011). However, these estimates are derived from relative sparse ground data sets that may have questionable sample selection properties; as a result, although landscape-level estimates may be unbiased, there is no evidence that these estimates are accurate and unbiased for smaller forest areas within the forest zone covered by EFI (Hayashi et al. 2015). The need to evaluate and potentially calibrate these estimates is required to assure appropriate forest management decisions are made on a local basis.

Localized bias can arise due to differences in past management treatments, microsites, and disturbance regimes, which can all produce substantial differences in stand structure and area-based forest attributes (Clutter et al. 1983; Oliver and Larson 1996). The high spatial resolution provided by EFI may connote an accuracy that is not realized on the ground. Evaluation and calibration of these estimates is required to make appropriate forest management decisions on a local scale. One solution is to locally calibrate EFI estimates using a sampling with covariates approach (Iles 2003, p. 443; Kershaw et al., 2016, pp. 343–353).

Sampling with covariates (Kershaw et al., 2016, pp. 343–353), double sampling, or “sampling to correct” (cf. Iles 2003, p. 443) is a widely used sampling approach. It is different than small area estimation where models are built for target areas with limited sample observations (Goerndt et al. 2011). With sampling with covariates, EFI estimates for the target area are used as the “large” sample and a “small” subsample of ground plots is obtained to correct EFI estimates. The sampling process is separated into two components. First is the sample selection method. In practice, systematic sampling is most common, since sample points can easily cover the target area (Freese 1962; Payandeh and Ek 1971; Iles 2003, pp. 171–182). However, applying variable probability selection methods, where selection probabilities are dependent on parameters of interest (e.g., EFI volume or biomass estimates), has potential to be a more efficient sampling strategy (Basu 1969). The second component is the sampling unit (plot type). Fixed area plots and horizontal point samples are common in

North America (Kershaw et al., 2016 Chap.9). Controlling the number of trees per plot results in trade-offs between increasing variability between plots (i.e., increasing required sample sizes) and decreasing time spent on each plot (Freese 1961; Gambill et al. 1985; Lynch 2017; Yang et al. 2017). Probabilistic subsampling for detailed tree measurements can improve sample efficiencies in many situations (Grosenbaugh 1979; Bell et al. 1983; Iles 2003, pp. 564–565; Marshall et al. 2004; Kershaw et al., 2016; Yang et al. 2017).

This paper explores the efficiency of different sample selection methods and different plot types on developing cost-effective and locally calibrated LiDAR-assisted inventories for small woodlots. The specific objectives include the following: (1) assessment of bias calibration; and (2) assessment of sampling efficiency, required sample sizes, and associated inventory costs of different sample selection method \times plot type combinations.

2 Material and methods

2.1 Study area

The study area was the 80-ha Femelschlag Research Site on the Noonan Research Forest (NRF; N 45°59'12", W 66°25'15"), managed by the University of New Brunswick's Faculty of Forestry and Environmental Management in New Brunswick, Canada. The NRF has a permanent grid of big BAF sample plots (BBAF) located on a North-South/East-West grid at 100-m spacing. The forest inventory is conducted every 10 years, with the inventory used in this study collected in 2012. The count BAF was 2 M (i.e., each tallied tree represented 2 m² ha⁻¹) and the measure BAF was 27 M. For each count tree, diameter at breast height (DBH, nearest 0.1 cm), status (live/dead), and species were measured and recorded. On the measure-trees, total height (HT, nearest 0.1 m) was measured. There were 1735 count trees and 100 measure-trees recorded on the 83 sample points contained within the Femelschlag Research Site.

The Femelschlag Research Site was established in 2014 with 83 fixed area plots (FAP) established on the same plot centers as the BBAF. The FAPs had a radius of 11.28 m (0.04 ha), and trees with ≥ 6 cm DBH were sampled. Tree species, status, DBH, and HT were recorded for 6252 sample trees. Table 1 summarizes the plots and tree measurements for both the BBAF and FAP data.

EFI estimates were provided by the Department of Energy and Resource Development, New Brunswick. EFI estimates were derived from ALS collected in 2015. The cell resolution was 20 m \times 20 m, and a total of 2181 cells covered the Femelschlag Research Site. While EFI provides estimates of several forest attributes for each EFI cell, only gross total

Table 1 Minimum, average, and maximum of number of trees per plot, number of species per plot, number of trees per ha (density), basal area (BA, m² ha⁻¹), diameter at breast height (DBH, cm), and total height (HT, m) for fixed area plots (FAP) and big BAF sample plots (BBAF) data

Parameter	FAP			BBAF		
	Min.	Mean	Max.	Min.	Mean	Max.
Number of trees (per plot)	23	75	175	12	21	31
Number of species (per plot)	2.0	5.4	8.0	2.0	5.2	8.0
Density (stems/ha)	575	1883	4375	457	1870	4233
BA (m ² ha ⁻¹)	21	40	54	24	42	62
DBH (cm)	6.0	14.5	69.0	6.0	22.9	79.9
HT (m)	3.1	12.0	33.4	–	–	–
Measure-trees (per plot)	–	–	–	0	1.2	5
Number of measure species (per plot)	–	–	–	0	1.0	3
Measure DBH (cm)	–	–	–	6.4	25.4	69.4
Measure HT (m)	–	–	–	3.6	16.4	24.5

volume (GTV, m³ ha⁻¹) was used in this study. The time differences between data collections and LiDAR estimates were ignored due to relatively slow tree growth (10–15 rings per cm) and minor natural or human disturbances in 2012–2015.

Operationally, aligning field samples with EFI cells will produce more accurate results. However, because our sampling grid and sample points were established prior to EFI, we did not have perfect spatial alignment with EFI cells. We assumed that our grid was a representative sample of the area around the sample locations. Thus, in the simulation, each EFI cell was assigned one of the 83 sample points based on distance between cell center and sample point location. This will add variability to our results but also will demonstrate the power of sampling with covariates.

2.2 Field sample point selection

In this study, the large sample was EFI cells and five sample selection methods were considered to select the small samples: (1) simple random sampling (SRS); (2) spatial systematic sampling (SYS); (3) systematic ordered-list sampling (SOL); (4) sampling with probability proportional to prediction (3P); and (5) list sampling (LIST).

SRS and SYS are two equal probability selection methods that are frequently used to select both large and/or small samples in double sampling. SYS is typically applied spatially, but can be applied to ordered lists of factors (Iles 2003, p. 176). The idea of SOL is to sort sample units from small to large by their covariate (i.e., GTV), then select sample units at a regular interval across the covariate range. Since SOL ensures equal distribution over the covariate range, it may provide a more efficient double sampling scheme than either SRS or SYS (Iles 2003, pp. 171–174).

3P and LIST are two variable probability selection methods. 3P was introduced by Grosenbaugh (1964, 1965).

Following Iles' (2003, pp. 450–451) procedure, a cruiser predicts (or estimates) an attribute related to the variable of interest for each individual in the population (tree, plot, ALS cell, etc.), and compares this prediction to a uniform random number drawn from the expected range of predictions. If the prediction \geq the random number, then the individual is measured. Although the sample size in 3P cannot be precisely controlled, Iles (2003) suggested controlling the range of random numbers to obtain the desired sample size. The maximum random number, KZ, is calculated using:

$$KZ = \frac{K}{p} = K \cdot \left(\frac{N}{n} \right) \quad (1)$$

where K = the expected value (mean) of the predictions; p = proportion of population to be sampled; N = population size; and n = desired sample size. In our application, since the EFI values are known and sample selection can be performed in advance, the number of units selected will be known in advance of field work despite being a random variable.

LIST is based on the probability proportional to contribution to the population total. In this procedure, as illustrated by Kershaw et al. (2016, pp. 353–356), the listed covariate (GTV, in this study) of sample units is firstly cumulated across the whole population. A sample unit is selected if a random number drawn between 0 and the total accumulation lies within the range of the cumulated sums. Thus, a sample unit that has a larger covariate will have higher probability of being selected and measured in detail. With LIST, the exact number of units selected is controlled and known in advance of field work.

2.3 Plot compilation

For each of the 83 sample points, FAP and BBAF data were available. To further explore the performance of tree subsampling within plots, four sample tree selection systems (plot types) were considered in this study: (1) fixed area plots with

all trees measured (FAP); (2) fixed area plots with 3P subsampling of measured trees (3PN); (3) horizontal point samples with all trees measured (HPS); and (4) big BAF sample plot (BBAF).

The 3PN was simulated using the FAP data. Height predictions, based on a height-diameter model (MacPhee et al. 2018), were compared with uniform random numbers between 0 and KZ (Eq. 1). If the predicted height \geq the random number, the measured tree height would be used. The desired tree sample size (n) was set at the same level of effort as observed in the big BAF samples ($n = 100$ trees across the 83 field plots). KZ (Eq. 1) was calculated using average predicted height ($K = 11.51$ m), and $N = 6252$ sample trees. Volume per ha ($\text{m}^3 \text{ha}^{-1}$) for a full FAP plot was calculated by estimating total volume for each tree using the measured DBH, HT, and Honer et al.'s (1983) metric volume equations, and then summing over all trees on the plot and multiplying by the expansion factor (25 for 11.28-m radius FAP). For the 3PN plots, the mean ratio of measured HT to predicted HT (\bar{R}_{HT}) for all measure trees was calculated and multiplied by each predicted HT, and the adjusted HTs and measured DBHs were used to estimate volume using the same process used on the full FAP.

To expand the BBAF data to represent HPS data, the NRF height-diameter model (MacPhee et al. 2018) was used to predict HT on all count trees. Volume was then estimated using Honer et al.'s (1983) metric volume equations and multiplied by each tree's associated expansion factor (varies with tree DBH) and summed to obtain volume per ha ($\text{m}^3 \text{ha}^{-1}$) for each sample point. For the BBAF, volume and volume to basal area ratio (VBAR) were calculated for each measured tree across the entire sample, and the mean VBAR was then multiplied by each plot's basal area per ha ($\text{m}^2 \text{ha}^{-1}$) to estimate volume per ha ($\text{m}^3 \text{ha}^{-1}$) for each sample point (Iles 2003, pp. 314–322, 564–565; Marshall et al. 2004). The EFI GTV values and the four different plot-level summaries are archived in the University of New Brunswick's DataVerse data archive (Hsu and Kershaw 2020).

2.4 Sample simulation

As described above, EFI cells formulate our large sample of covariates and we will select a smaller subsample of these cells to obtain simulated field measurements. Sample sizes between 10 and 50, in steps of 5, were simulated for each of the five sample selection methods and each plot type was used at each selected cell. Each sample selection method \times sample size was simulated 100 times. Sampling was conducted without replacement except for LIST which, as typically implemented, was conducted with replacement (Freese 1962; Kershaw et al., 2016, pp. 353–356). Preliminary analyses showed that the GTV for our study area did not have

appreciable bias (approximately 2–9%, depending on plot type) relative to the field plots (bias = cell GTV – Field Volume). To test our hypothesis that sampling with covariates and ratio estimation can be used to correct bias in EFI cells and to provide meaningful comparisons between different sample design combinations, we added multiplicative biases into each EFI cell by generating random numbers between 1.05 and 1.15 for 10% bias and 1.15 and 1.25 for 20% bias (the adjustment was in one direction because we were testing effects of bias, not error). All sample simulations were carried out in the R statistical program (R Development Core Team 2019).

2.5 Sample compilation

In this study, ratio estimation (Freese 1962; Iles 2003, pp. 314–322; Kershaw et al., 2016, pp. 351–353) was used to calibrate (or correct) GTV using the small sample field volume estimates (VOL). Ratio estimation can be considered within either a design-based or a model-based framework (e.g., Gregoire 1998; Magnussen 2015). Here, we choose the model-based (or model-dependent) approach. Because GTV is expected to be very similar to VOL, the relationship between GTV and VOL should be linear and pass through 0. For the i th selected EFI cell in the j th replicate sample in the k th sample design combination (sample selection method \times plot type \times sample size), the VOL to GTV ratio (R_{ijk}) was calculated using:

$$R_{ijk} = \frac{\text{VOL}_{ijk}}{\text{GTV}_{ijk}} \quad (2)$$

where VOL_{ijk} = field estimated volume using one of the four plot types; and GTV_{ijk} = the GTV of selected EFI cell. The mean ratio (\bar{R}_{jk}) was then calculated for the j th replicate sample in the k th sample design combination:

$$\bar{R}_{jk} = \frac{\sum_{i=1}^m R_{ijk}}{m_{jk}} \quad (3)$$

where m_{jk} = the number of samples. We note that under the design-based approach, different estimators of the population mean would be necessary, depending on the selection probabilities of the individual cells under the design (Thompson 2012 Chap. 7). From a model-based perspective, the sampling probabilities are irrelevant so long as selection is “noninformative” in that it is not influenced by the residual value itself (e.g., Chambers and Clark 2012); a single estimating equation can be used, depending on the characteristics of the error term for individual cells. For further discussion on noninformative samples in forest inventory, see Magnussen (2015). Equation 3, a mean-of-ratios estimator, is the minimum-variance model-unbiased estimator when the residual variance is proportional to GTV^2 , or the residual standard

deviation is proportional to GTV (Brewer 1963). Adjusted GTV (VOL_{ijk}) was then obtained using:

$$VOL_{ijk} = \bar{R}_{jk} \cdot GTV_i \quad (4)$$

The mean adjusted GTV (\overline{VOL}_{jk}) was then calculated as:

$$\overline{VOL}_{jk} = \frac{\sum_{i=1}^N VOL_{ijk}}{N} \quad (5)$$

The overall error of the sample can be considered from two different perspectives. If we view the LiDAR-derived GTV as a random variable associated with its own standard error, then overall sample error can be calculated using Bruce's formula (Iles 2003, p. 588; Marshall et al. 2004; Kershaw et al., 2016, p. 380) as derived from Goodman (1960):

$$se\%(\overline{VOL}_{jk}) = \sqrt{se\%(\bar{R}_{jk})^2 + se\%(GTV)^2} \quad (6)$$

where $se\%(\overline{VOL}_{jk})$ = the percent standard error of the mean adjusted GTV; $se\%(\bar{R}_{jk})$ = the percent standard error of the mean ratio; and $se\%(GTV)$ = the percent standard error of the mean original GTV. The percent standard error was calculated using:

$$se\%(\bar{X}) = 100 \cdot \left(\frac{s(X)/\sqrt{m}}{\bar{X}} \right) = \left(\frac{CV(X)}{\sqrt{m}} \right) \quad (7)$$

where $X = R$ or GTV ; $\bar{X} = \bar{R}$ or \overline{GTV} (averages); $s(X)$ = standard deviation of X , m = sample size; and $CV(X)$ = coefficient of variation of X . The use of Eq. 7 for R conforms to the usual treatment of the plots as independent estimates. However, for GTV , one might argue that ignoring spatial autocorrelation leads the resulting $se\%(GTV)$ to be a slight underestimate. On the other hand, a strong case can be made that the LiDAR-derived GTV , for a given property and a given LiDAR collection, is a fixed value and that randomness only enters through variation in the VOL_{ijk} of the individual sampled cells. In that case, the term $se\%(GTV)$ in Eq. 6. equals zero, and spatial autocorrelation becomes a non-issue. Ignoring spatial autocorrelation, $se\%(GTV)$ calculated using Eq. 7 is about two orders of magnitude smaller than $se\%(\bar{R}_{jk})$ in all cases studied, such that $se\%(\overline{VOL}_{jk})$ is identical to within 5 or 6 decimal places. Even with a generous allowance for spatial autocorrelation (e.g., an increase of an order of magnitude), the influence of $se\%(GTV)$ on $se\%(\overline{VOL}_{jk})$ is negligible and can hardly be represented meaningfully or discerned in graphical or tabular presentation of the results; it is certainly too small to be of management importance, so we consider it no further and use Bruce's formula (Eq. 6) in all analyses presented here.

2.6 Data analyses

2.6.1 Sample design effects

Effects of different sample design combinations on the distributions of \overline{VOL}_{jk} and $se\%(\overline{VOL}_{jk})$ were examined using horizontal error bar plots and bean plots (Kampstra 2008), respectively. To assess overall sampling efficiency and determine the variance contributions of selection methods and plot types, a nonlinear mixed effects model was fitted:

$$se\%(\overline{VOL}_{jk}) = \frac{b_0}{m^{b_1}} \quad (8)$$

where m = sample size; and b_i = regression coefficients. Random effects were fitted using sample selection method and plot type nested within sample selection method to both coefficients. With the use of Eq. 8, minimum sample size requirements were determined for the 5% error level, and the design effect (Särndal et al. 1992) was calculated using:

$$DE_{st} = \frac{m_{5\%,s}}{m_{5\%,t}} \quad (9)$$

where DE_{st} = the design effect between sample design s and t ; and $m_{5\%,i}$ = the sample size requirement to achieve 5% error from sample design s or t .

2.6.2 Influence of cost constraints on optimal sample design

A field survey design usually needs to consider both statistical efficiency and time or cost (Smith 1938). Both Lynch (2017) and Yang et al. (2017) showed that optimal sample design varies in relationship to cost constraints. Building upon the work of Lynch (2017), Yang et al. (2017) expressed total inventory cost as the sum of overhead costs (e.g., the planning, preparation and compilation), total measurement costs (C_{Meas}), and plot-to-plot travel costs (C_{Travel}). Overhead costs are fixed. C_{Meas} is composed of plot establishment costs, tree identification costs, and tree measurement costs:

$$C_{Meas} = C_P + k \cdot E[c] \cdot m + r \cdot M \quad (10)$$

where C_P = sample point establishment costs; k = per tree costs to determine an "in" tree, identify species, and measure DBH; $E[c]$ = expected number of trees per sample point; m = required sample size; r = cost to measure other tree attributes (HT, in this study); and M = the total number of measure trees. Zeide (1980) estimated travel distance between two sampling points using a square grid: $d = \sqrt{A/m}$ (A = total area of forest in square of units of distance). The equation for C_{Travel} can then be expressed as:

$$C_{Travel} = C_d \cdot m \cdot \sqrt{A/m} = C_d \cdot \sqrt{A \cdot m} \quad (11)$$

where C_d = travel cost per meter.

Given Eqs. 10 and 11, costs for a given sample size and sample design were estimated. Finally, the design effect was expressed as a ratio of costs:

$$DE_{st} = \frac{C_{5\%,s}}{C_{5\%,t}} \quad (12)$$

The inventory costs used in this study were based on ancillary data collected on the NRF (Yang et al. 2017): $C_0 = 1300$, $C_p = 2.70$, $k = 0.675$, $r = 4.41$, and $C_d = 0.0945$ (2015 Canadian dollars). Crew size was assumed constant (two people) for all designs.

3 Results

Differences in mean adjusted volumes \overline{VOL}_{jk} and associated standard errors across the three bias levels (no bias, 10%, and 20% added bias) were negligible, except for SRS. It is important to note the negligible differences demonstrate the efficiency of sampling with covariates to correct for local bias. However, because of the small differences in results, we only present results based on 10% bias to illustrate the ability of the ratio estimation and the efficiency of sampling with covariates.

3.1 Sample design effects

Figure 1 shows the mean and range of estimated \overline{VOL}_{jk} by different sample designs at 10% added bias. Field measured volume varied by plot type (Fig. 1): FAP averaged 272.8 (± 6.69) $m^3 ha^{-1}$; 3PN averaged 270.5 (± 5.78) $m^3 ha^{-1}$; HPS averaged 247.0 (± 7.97) $m^3 ha^{-1}$; and BBAF averaged 256.2 (± 7.21) $m^3 ha^{-1}$. Although the range of \overline{VOL}_{jk} for equal probability selections was greater than the 95% confidence interval (CI) of 83 field plot measured volumes and greater than variable probability selections, mean \overline{VOL}_{jk} for every combination was close to the mean field measured volume (Fig. 1), except for SOL. As sample size increased, 3P and LIST generally produced narrower ranges in estimated \overline{VOL}_{jk} than equal probability selection methods (Fig. 1). SOL consistently had the most extreme \overline{VOL}_{jk} when sample size was 10 (Fig. 1a); however, this difference was minimal as sample size increased. 3P generally had the smallest range of estimated \overline{VOL}_{jk} for most plot types and sample sizes, except LIST had smaller ranges than 3P for HPS and BBAF plots for sample size 20 (Fig. 1b).

The $se\%(\overline{VOL}_{jk})$ decreased with increasing sample size for all sample design combinations (Fig. 2). While all selection methods produced many sample instances with low $se\%(\overline{VOL}_{jk})$, the equal probability methods often produce extreme values which increased their mean $se\%(\overline{VOL}_{jk})$. 3P

and LIST had lower mean $se\%(\overline{VOL}_{jk})$ than all equal probability selection methods. Plots using subsampling methods (3PN and BBAF) had smaller mean $se\%(\overline{VOL}_{jk})$ than plots where all trees were measured (FAP and HPS; Fig. 2).

The fixed effects component of Eq. 8 only explained 13% of the variation in $se\%(\overline{VOL}_{jk})$ and the full model including the random effects explained 19% with an overall root mean square error of 3.5%. The fixed parameter estimates were as follows: $b_0 = 25.0915$ (1.98981) and $b_1 = 0.4223$ (0.02181). The standard errors for the random effects associated with sample selection method were as follows: $s(b_0) = 4.2605$ and $s(b_1) = 0.0431$; and those associated with plot type nested within sample selection method were as follows: $s(b_0) = 3.0578 \times 10^{-8}$ and $s(b_1) = 3.1263 \times 10^{-2}$. While the model accounts for a small percentage of the variation due to high variability in some sample designs (especially in SOL), the results are still informative and useful for sample design comparisons.

Table 2 shows the combined fixed + random effects coefficient estimates by sample selection methods and plot types. The b_0 is the theoretical population standard deviation, and the b_1 represents the rate of standard error reduction with increasing sample size. Larger b_1 represent more efficient designs while larger b_0 represent more variable designs. Theoretically, b_1 should equal 0.5. The b_0 of SRS and SYS were similar; SOL had the largest b_0 , and 3P had the lowest. The b_0 of LIST was higher than SRS and SYS. The b_1 for SRS and SYS were usually lower than the ones obtained for variable probability selection methods. The b_1 were slightly lower than 0.5 in general, except for LIST with FAP and 3PN. Based on the same sample selection method, the use of subsampling plot types (3PN and BBAF) also had higher b_1 than all trees measured within plots (FAP and HPS).

Table 3 shows the required sample sizes for 5% standard error derived from Eq. 8. The required sample sizes did not increase substantially across the different bias levels, except for SRS. For SRS, the required sample sizes increased proportionally to the bias level increment (Table 3). LIST required the lowest sample sizes for all plot types and bias levels. Figure 3 shows the design effect based on minimum required sample sizes. FAP under SRS and SYS were more efficient than HPS and BBAF, except when these plot types were selected using 3P or LIST. 3P and LIST for all plot types were generally more efficient than other selection methods, especially for the two fixed area plot types (FAP and 3PN). LIST with 3PN was more efficient than all other sample selection method \times plot type combinations.

3.2 Costs

Unlike the higher sample size design effects (i.e., more sample efficiency) associated with FAP (Fig. 3), when costs were

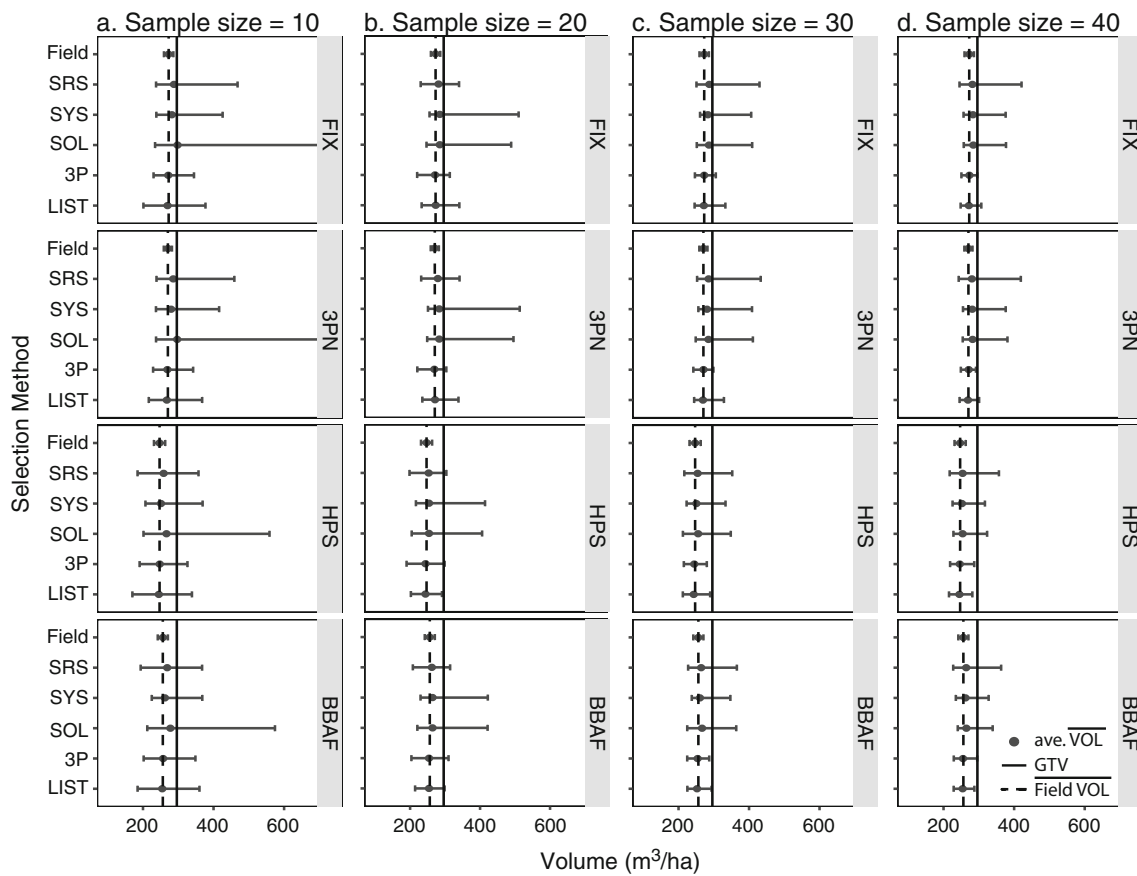


Fig. 1 Comparison of field measured volumes with 95% confidence intervals based on the full 83 samples and the mean and range of VOL (with 10% bias added) for the 100 replicates by five sample selection methods and four plot types. (Selection method and plot type acronyms

are as defined in Table 2; $ave.VOL$ = mean of the 100 VOL s, GTV = mean of uncorrected EFI volume with 10% bias, $Field.VOL$ = mean of field measured volume

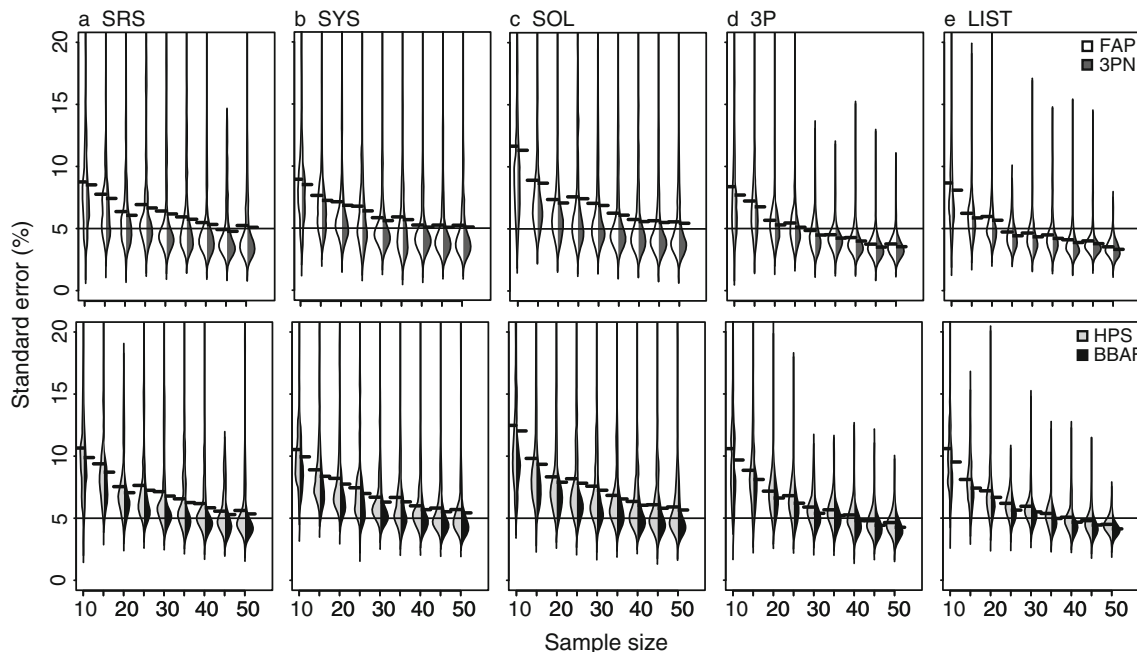


Fig. 2 The averages and distributions of $se\%(VOL)$ with 10% bias by five sample selection methods and four plot types across the range of sample sizes (10 to 50). (Selection method and plot type acronyms are as defined in Table 2)

Table 2 Coefficient estimates (fixed + random effects) of the nonlinear mixed effect model (Eq. 8) for estimating minimum sample size requirements for the scenario with 10% added bias by selection method and plot type. (Sample selection methods: *SRS* = simple random sampling, *SYS* = systematic sampling, *SOL* = systematic ordered-list sampling, *3P* = sampling with probability proportional to prediction, *LIST* = list sampling. Plot types: *FAP* = fixed area plot, *3PN* = fixed area plot with 3P subsampling of measured trees, *HPS* = horizontal point sample, *BBAF* = big BAF sampling)

Selection method	b_0	b_1 (Plot type)			
		FAP	3PN	HPS	BBAF
SRS	22.4523	0.3867	0.3976	0.3445	0.3629
SYS	22.5925	0.3879	0.4011	0.3480	0.3654
SOL	32.8344	0.4657	0.4741	0.4395	0.4536
3P	21.9240	0.4479	0.4686	0.3780	0.4043
LIST	25.6546	0.5039	0.5238	0.4337	0.4600

considered, FAP no longer had any positive design effects except when used with variable probability selection methods (Fig. 4). As with sample size (Fig. 3), variable probability selection was more efficient than equal probability designs (Fig. 4). Subsampling plot types had larger cost-based design effects than the plot types with all trees measured (Fig. 4). As a result, the total costs associated with 3PN and BBAF were much less than FAP and HPS (Fig. 5). Because FAP had more expected “in” trees per plot than HPS (Table 3), even though 3PN had lower required sample sizes than BBAF (Table 3, Fig. 3), the costs for 3PN were still higher than BBAF (Figs. 4 and 5).

4 Discussion

In forestry, inventory errors in the 10–20% range are common and acceptable, while the estimates should be unbiased (Schreuder et al. 1993; Gregoire and Valentine 2008). Under the approach of sampling with covariates and ratio estimation,

auxiliary variables which support a sampling process should not be a source of bias (Iles 2003) as long as the primary variable (small sample) is unbiased. Typically, bias is less than 5%, and biases in our corrected results ranged from 0.1 to 5.0% (except SOL which bias was around 8% in small sample size). As a result, bias can generally be corrected using ratio estimation without substantially increasing sampling effort. Our results support this opinion, except for SRS (Table 3). Because SRS does not consider any information about the population and the added bias increased population variance, SRS required more samples as bias increased (Table 3). *SYS* and *SOL*, which also are equal probability selection methods, ensure spatial or distributional coverage across the population range, and, as a result, the added bias did not have a substantial impact on sample sizes (Table 3).

In theory, *SOL* should be more efficient than *SRS* and *SYS* because *SOL* samples across the population range (Iles 2003, pp. 175–183). However, in our case, *SOL* produced the most extreme volume estimates (Fig. 1) and required more samples to achieve a given standard error (Table 3, Fig. 2). It is because the ratios of field volume to EFI volume were calculated using the nearest field plot and the selected EFI cells were not spatially aligned. A large-valued field plot, coupled with an EFI cell centered on a small gap or at an edge, can result in a very large ratio, which could bias the overall average ratio and subsequent corrected volume estimate. Given that *SOL* ensures sampling across the range of EFI values, these extreme estimates were more likely due to our data limitations, especially at the lowest sample sizes (Fig. 1a). At the larger sample sizes (Fig. 1c, d), *SOL* was on par with *SYS* and slightly better than *SRS*.

Most ALS inventory studies align field plots and ALS cells (White et al. 2013). This was not done here because the field data sets were established independently of EFI cells. We used all EFI cells and matched closest field plots, which was purposely done to increase population size (large sample) and variability for our sample simulations. Despite this, the sampling with covariates and ratio estimation approach was still robust, and produced unbiased estimates with relatively small,

Table 3 The required sample sizes under 5% sampling error for each combination of sample selection methods and plot types by three bias levels (No bias, 10% bias, and 20% bias). (Selection method and plot type acronyms are defined in Table 2)

Plot type	# trees/plot (Total/ meas)	Selection method														
		SRS			SYS			SOL			3P			LIST		
		No bias	10%	20%	No bias	10%	20%	No bias	10%	20%	No bias	10%	20%	No bias	10%	20%
FAP	75.3/75.3	43	48	52	49	49	50	57	57	58	27	28	29	25	26	27
3PN	75.3/1.3	39	44	47	43	43	44	53	53	55	23	24	25	22	23	24
HPS	20.9/20.9	61	79	84	76	77	78	72	73	75	47	50	46	44	44	42
BBAF	20.9/1.2	52	63	67	62	63	63	64	64	65	37	39	37	35	35	35

*Total = total number of trees per plot, Meas = number of measured trees per plot

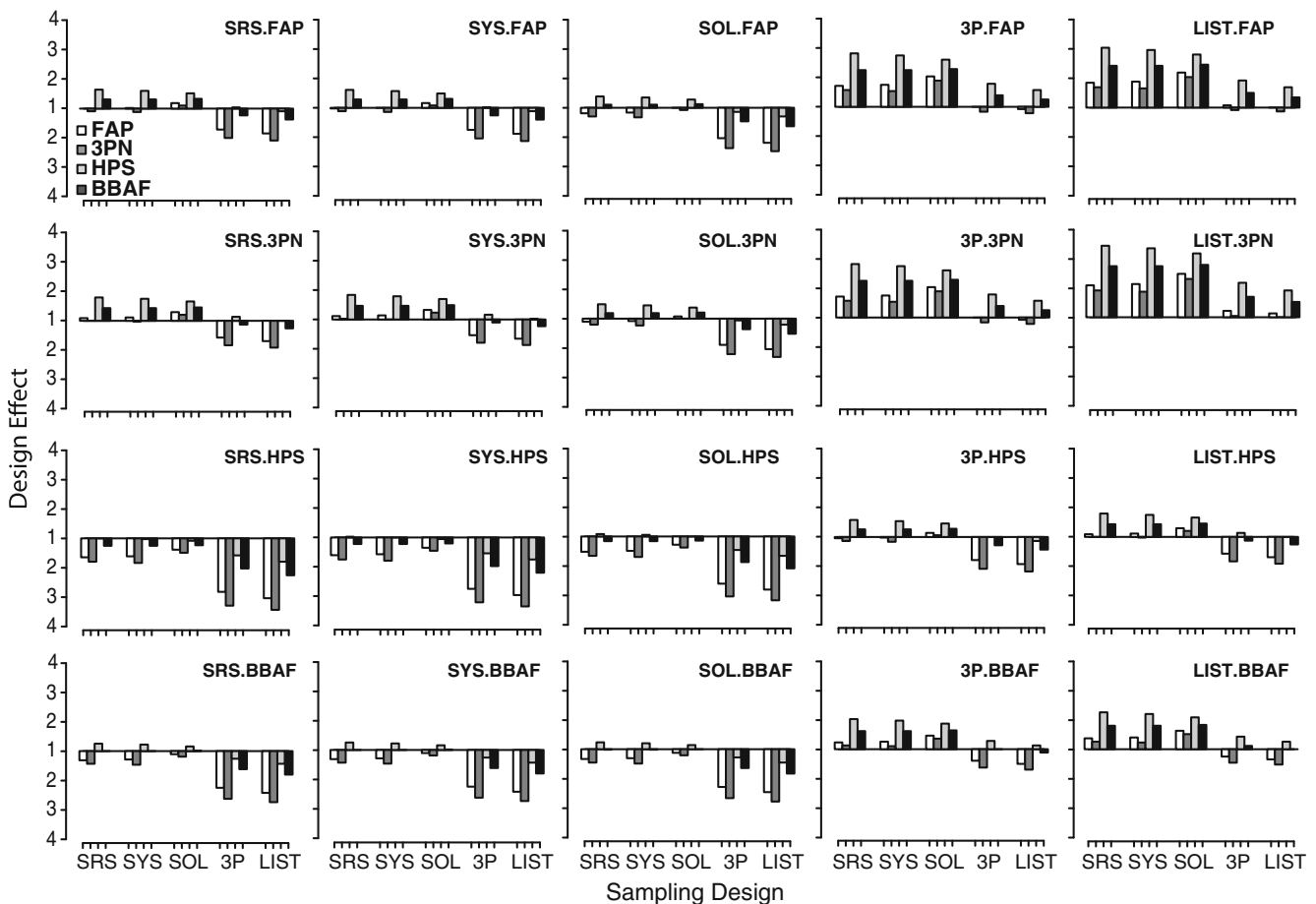


Fig. 3 Sample size–based design effect comparisons by five sample selection methods and four plot types. Sample size calculations were based on a desired standard error of 5%. Upward pointing (positive) bars indicate that the named panel sample design is more effective than the x-axis

sample design ($DE = x\text{-axis sample size}/\text{panel sample size}$). Downward pointing (negative) bars indicate the x-axis design is more efficient than the named panel design ($DE = \text{panel design sample size}/x\text{-axis sample size}$). (Selection method and plot type acronyms are as defined in Table 2)

cost-effective sample sizes. This study demonstrates that it is possible to use existing ground data that does not exactly align directly with EFI cells for the purpose of locally calibrating LiDAR-assisted predictions. The fundamental outcome of this study is the requirement for an underlying, probabilistic sample at the second stage (Iles 2003, pp. 314–322). The small sample enables us to estimate a total, and the large sample enables us to spread that total over the area of interest (Freese 1962; Iles 2003, pp. 314–322).

Variable probability selection, which allows sampling with probability proportional to the variable of interest (Basu 1969; Iles 2003 Chap.11), minimized extreme value estimates (Fig. 1), reduced overall sample size requirements (Table 3, Figs. 2 and 3), and reduced inventory costs (Figs. 4 and 5). While West (2017) found that variable probability selection was not superior to SRS, this study shows both 3P and LIST had better performance than any of the equal probability selection methods for all criteria evaluated, similar to the results in Yang et al.'s (2019) study. It should be noted that West (2017) only implemented 3P and a variant of 3P that did not

require a census of covariates. In this study, LIST typically outperformed 3P.

In traditional forest inventory, LIST and 3P methods were not widely implemented because of the census demands (either a priori for LIST or concurrent for 3P). As shown here, estimates derived from ALS provide suitable lists of covariates for LIST and 3P. Besides the EFI estimated volume used in this study, covariates extracted directly from ALS point cloud data (e.g., LiDAR metrics) also can be used (Yang et al. 2019). Similar to Yang et al.'s (2019) work, Parker and Evans (2004, 2009) showed that double sampling could be an effective approach to derive LiDAR estimates for forest and individual tree attributes, but they only used systematic selection. Other authors have explored additional sample selection criteria such as space filling designs (Junttila et al. 2013) and local pivotal methods (LPM; Grafström et al. 2012; Grafström and Ringvall 2013). These designs are similar to SOL in that they attempt to balance sample allocation across multiple spatial and statistical dimensions. Yang et al.'s (2019) results found that LPM performed as well as 3P and LIST when

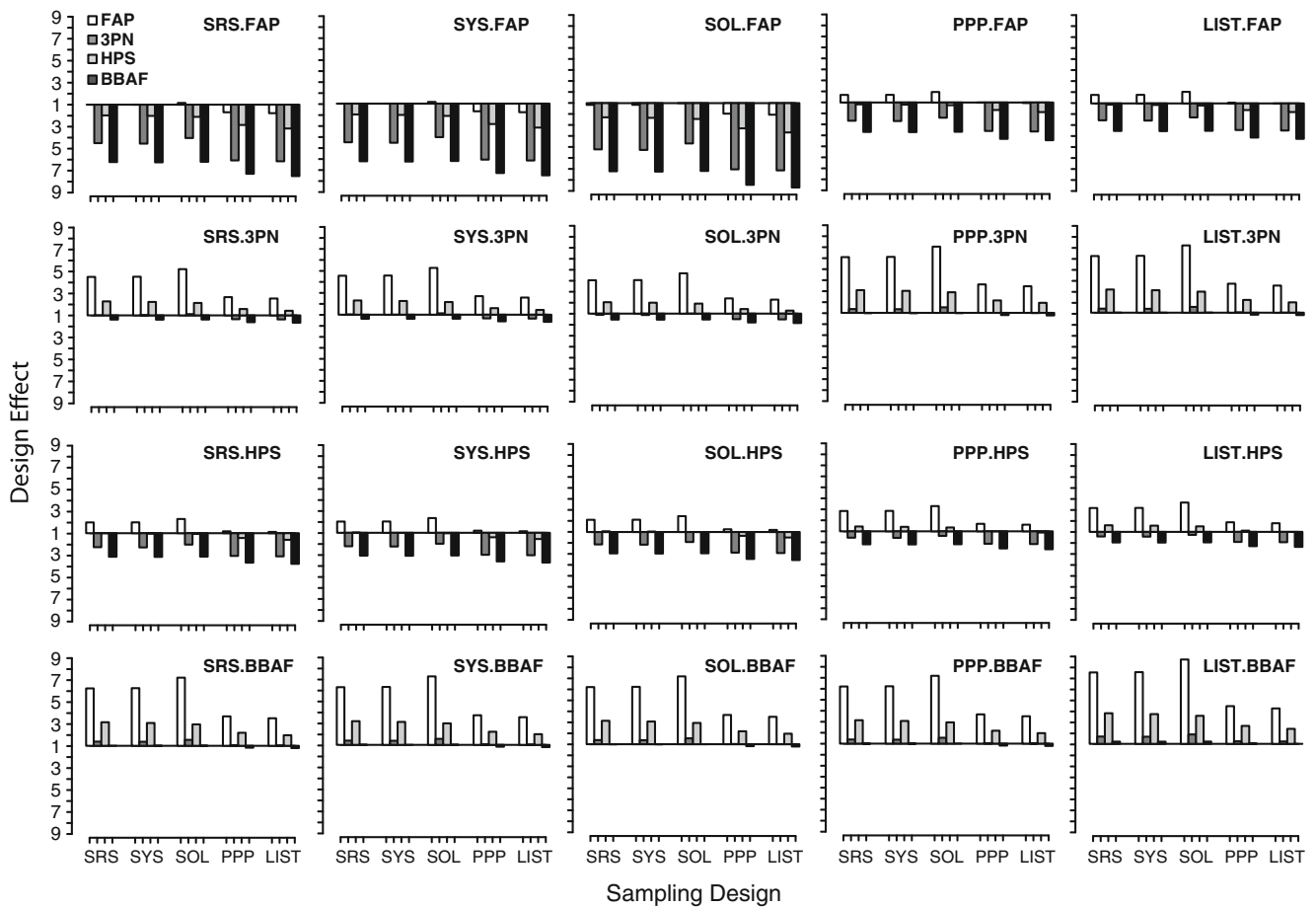
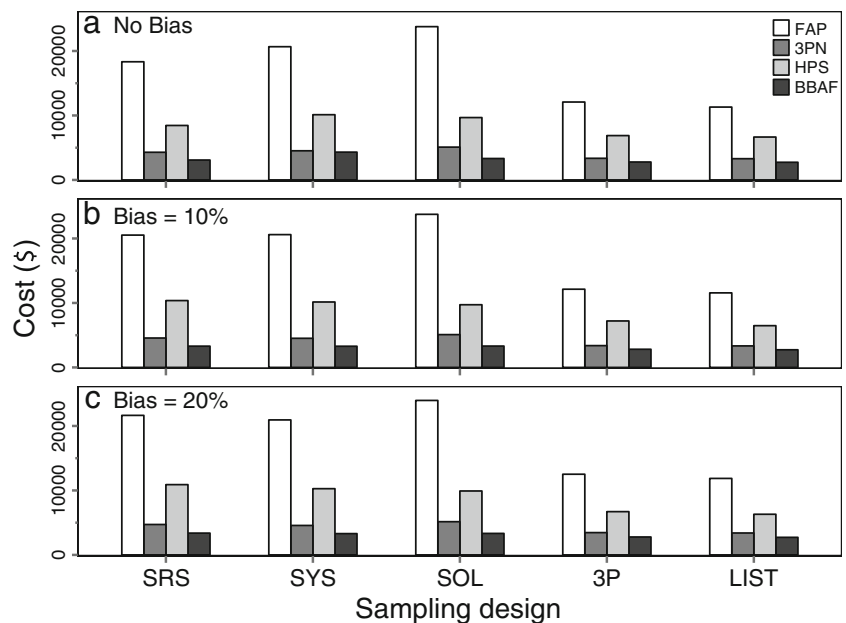


Fig. 4 Cost-based design effect comparisons by five sample selection methods and four plot types. Sample size calculations were based on a desired standard error of 5%. Upward pointing (positive) bars indicate that the named panel sampling design is more cost-effective than the x-

axis sample design ($DE = x\text{-axis cost}/\text{panel cost}$). Downward pointing (negative) bars indicate the x-axis design is more cost-efficient than the named panel design ($DE = \text{panel design cost}/x\text{-axis cost}$). (Selection method and plot type acronyms are as defined in Table 2)

Fig. 5 Total costs by five sample selection methods and four plot types across three bias levels (No bias; 10% bias, and 20% bias). (Selection method and plot type acronyms are as defined in Table 2)



developing forest-level LiDAR-assisted inventories. As spatial scale increases, stratified sampling or cluster sampling may be more effective (Schreuder et al. 1993; Gregoire and Valentine 2008). However, even with these designs, the principles of variable probability selection can be considered to select sampling units.

The inventory cost of a sample procedure is influenced by two factors: (1) required sample size, which is mainly influenced by sample selection methods (Table 3, Figs. 2 and 3); and (2) the number of “in” trees and measure-trees, which are dependent on plot types. Sample size generally is the main factor used to assess sample efficiency. However, because the spatial extent in this study was only 80 ha, which can be considered as a stand-level or small woodlot inventory, the cost for tree measurement is much higher than for traveling between plots (Yang et al. 2017). Thus, plot type was the major factor impacting total costs rather than selection methods. HPS, as a variable probability selection method, had lower numbers of “in” trees than FAP (Table 3). Although in our study area, HPS and BBAF required larger sample sizes than FAP and 3PN, they still had lower costs. Plot types with subsampling saved much effort in measuring trees (Table 3). Ratios used in 3PN and BBAF were effective for reducing variation between individual measurements (Iles 2003, pp. 564–565; Marshall et al. 2004).

5 Conclusion

Our study shows how LiDAR-assisted forest inventory estimates can be corrected for localized conditions, and the potential of using sampling with covariates and ratio estimation. This approach can be applied to not only LiDAR-assisted estimates but to other remote sensing techniques, such as satellite data and even ground-based optical sensors (Hsu 2019). And it can be used to estimate any parameter of interest, including biomass or carbon content (Chen et al. 2020). As ALS point cloud data and EFI estimates are becoming publicly available (especially in Atlantic Canada), the applications presented in this study become feasible for both large industrial forest landholders and smaller private woodlot owners, giving them high-resolution forest information that is locally calibrated to support management decisions.

Acknowledgments We are grateful to the dedicated work of the 2014 summer field crew: Mike Hutchinson, Shawn Donovan, Charles MacPhee, and Dan Kilham.

Funding information This study was funded by Natural Sciences and Engineering Research Council of Canada (Discovery Grant program) (RGPIN04280), and the New Brunswick Innovation Foundation Research Assistantship Initiative (RAI2018-025).

Data availability The datasets generated during and/or analyzed during the current study are available in the University of New Brunswick’s DataVerse data archive, <https://doi.org/10.25545/Z8WRBJ>.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Andersen H-E, Strunk JL, Temesgen H, Atwood D, Winterberger K (2011) Using multilevel remote sensing and ground data to estimate forest biomass resources in remote regions: a case study in the boreal forests of interior Alaska. *Can J Remote Sens* 37:596–611. <https://doi.org/10.5589/m12-003>
- Asner GP, Hughes RF, Mascaro J et al (2011) High-resolution carbon mapping on the million-hectare island of Hawaii. *Front Ecol Environ* (e-View) 28. <https://doi.org/10.1890/100179>
- Basu D (1969) Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā: The Indian Journal of Statistics* 31:441–454
- Bell JF, Iles K, Marshall D (1983) Balancing the ratio of tree count-only sample points and VBAR measurements in variable plot sampling. In: Bell JF, Atterbury T (eds) *Proceedings: renewable resource inventories for monitoring changes and trends*. College of Forestry, Oregon State University, Corvallis, OR, pp 699–702
- Brewer KRW (1963) Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics* 5:93–105. <https://doi.org/10.1111/j.1467-842X.1963.tb00288.x>
- Canada NR (2013) Enhanced forest inventory techniques. <https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/sustainable-forest-management/measuring-reporting/forest-inventory/enhanced-forest-inventory-techniques/13421>. Accessed 17 Aug 2019
- Chambers RL, Clark RG (2012) *An introduction to model-based survey sampling with applications*. Oxford University Press, Oxford ; New York
- Chen Y, Kershaw JA, Hsu Y-H, Yang T-R (2020) Carbon estimation using sampling to correct LiDAR-assisted enhanced forest inventory estimates. *For Chron* 96:9–19. <https://doi.org/10.5558/fc2020-003>
- Clutter JL, Fortson JC, Pienaar LV, et al (1983) *Timber management. A quantitative approach, first*. John Wiley & Sons, New York
- Corona P, Fattorini L (2008) Area-based lidar-assisted estimation of forest standing volume. *Can J For Res* 38:2911–2916. <https://doi.org/10.1139/X08-122>
- Dick A (2019) Enhanced forest inventory (EFI) adoption in New Brunswick: progress to date and future directions. <http://www.cif-ifc.org/wp-content/uploads/2018/12/Enhanced-Forest-Inventory-EFI-Adoption-in-NB.pdf>
- Freese F (1961) Relation of plot size to variability: an approximation. *J For* 59:679
- Freese F (1962) *Elementary forest sampling*. US Department of Agriculture
- Gambill CW, Wiart HV, Yandle (1985) Optimum plot size and BAF. *For* 31:587–594
- Goerndt ME, Monleon VJ, Temesgen H (2011) A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Can J For Res* 41:1189–1201. <https://doi.org/10.1139/x11-033>
- Goodman LA (1960) On the exact variance of products. *J Am Stat Assoc* 55:708–713

- Grafström A, Ringvall AH (2013) Improving forest field inventories by using remote sensing data in novel sampling designs. *Can J For Res* 43:1015–1022. <https://doi.org/10.1139/cjfr-2013-0123>
- Grafström A, Lundström NLP, Schelin L (2012) Spatially balanced sampling through the pivotal method. *Biometrics* 68:514–520. <https://doi.org/10.1111/j.1541-0420.2011.01699.x>
- Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. *Can J For Res* 28:1429–1447. <https://doi.org/10.1139/x98-166>
- Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. Chapman Hall/CRC Press
- Grosenbaugh LR (1964) Some suggestions for better sample–tree measurement. In: Proceedings of the 1963 Society of American Foresters National Convention. Society of American Foresters, pp 36–42
- Grosenbaugh LR (1965) THREE-PEE SAMPLING THEORY and program “THRP” for computer generation of selection criteria. USDA Forest Service, RP-PSW-021, Pacific Southwest Forest Experiment Station
- Grosenbaugh LR (1979) 3P sampling theory, examples, and rationale. US Department of Interior, Bureau of Land Management
- Hayashi R, Kershaw JA Jr, Weiskittel AR (2015) Evaluation of alternative methods for using LiDAR to predict aboveground biomass in mixed species and structurally complex forests in northeastern North America. *Mathematical and Computational Forestry and Natural Resources Sciences* 7:49–65
- Honer TG, Ker MF, Alemdag IS (1983) Metric timber tables for the commercial tree species of central and eastern Canada. Canadian Forestry Service, M-X-140, Maritimes Forest research Centre
- Hsu Y-H (2019) Applications of variable probability sampling using remotely sensed covariates. University of New Brunswick, Master Thesis
- Hsu Y-H, Kershaw JA (2020) Acadian Forest volume estimates derived from airborne LiDAR, big BAF sample plots, and fixed area plots on the Noonan Research Forest. University of New Brunswick. Dataverse Research Data Repository. [Dataset]. V1. <https://doi.org/10.25545/Z8WRBj>
- Iles K (2003) A sampler of inventory topics, 2nd edn. Kim Iles and Associates, Nanaimo, BC
- Junttila V, Finley AO, Bradford JB, Kauranne T (2013) Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. *For Ecol Manag* 292:75–85. <https://doi.org/10.1016/j.foreco.2012.12.019>
- Kampstra P (2008) Beanplot: a boxplot alternative for visual comparison of distributions. *J Stat Soft* 28:1–9. <https://doi.org/10.18637/jss.v028.c01>
- Kershaw JA Jr, Ducey MJ, Beers TW, Husch B (2016) Forest Mensuration, 5th edn. Wiley/Blackwell, Hoboken
- Lynch TB (2017) Optimal plot size or point sample factor for a fixed total cost using the Fairfield Smith relation of plot size to variance. *Forestry* 90:211–218. <https://doi.org/10.1093/forestry/cpw038>
- MacPhee C, Kershaw JA Jr, Weiskittel AR et al (2018) Comparison of approaches for estimating individual tree height–diameter relationships in the Acadian Forest region. *Forestry* 91:132–146. <https://doi.org/10.1093/forestry/cpx039>
- Magnussen S (2015) Arguments for a model-dependent inference? *Forestry* 88:317–325. <https://doi.org/10.1093/forestry/cpv002>
- Marshall DD, Iles K, Bell JF (2004) Using a large-angle gauge to select trees for measurement in variable plot sampling. *Can J For Res* 34: 840–845. <https://doi.org/10.1139/X03-240>
- Næsset E (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens Environ* 80:88–99. [https://doi.org/10.1016/S0034-4257\(01\)00290-5](https://doi.org/10.1016/S0034-4257(01)00290-5)
- Oliver CD, Larson BC (1996) Forest stand dynamics, updated. Wiley, New York
- Parker RC, Evans DL (2004) An application of LiDAR in a double-sample forest inventory. *W J Appl For* 19:95–101. <https://doi.org/10.1093/wjaf/19.2.95>
- Parker RC, Evans DL (2009) LiDAR Forest inventory with single-tree, double-, and single-phase procedures. *Int J For Res* 2009:6–6. <https://doi.org/10.1155/2009/864108>
- Payandeh B, Ek AR (1971) Observations on spatial distribution and the relative precision of systematic sampling. *Can J For Res* 1:216–222. <https://doi.org/10.1139/x71-029>
- R Development Core Team (2019) R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Särndal C-E, Swensson B, Wretman JH (1992) Model assisted survey sampling, 1. softcover print. Springer, New York
- Schreuder HT, Gregoire TG, Wood GB (1993) Sampling methods for multiresource forest inventory, first. John Wiley & Sons, New York
- Smith HF (1938) An empirical law describing heterogeneity in the yields of agricultural crops. *J Agric Sci* 28:1–23. <https://doi.org/10.1017/S0021859600050516>
- Thompson SK (2012) Sampling, 3rd edn. Wiley, New York
- West PW (2017) Simulation studies to examine bias and precision of some estimators that use auxiliary information in design-based sampling in forest inventory. *NZ j of For Sci* 47:22. <https://doi.org/10.1186/s40490-017-0101-7>
- White JC, Wulder MA, Varhola A, Vastaranta M (2013) A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach (Version 2.0). Natural Resources Canada - Canadian Forest Service
- Yang T-R, Hsu Y-H, Kershaw JA Jr et al (2017) Big BAF sampling in mixed species forest structures of northeastern North America: influence of count and measure BAF under cost constraints. *Forestry* 90:649–660. <https://doi.org/10.1093/forestry/cpx020>
- Yang T-R, Kershaw JA, Weiskittel AR et al (2019) Influence of sample selection method and estimation technique on sample size requirements for wall-to-wall estimation of volume using airborne LiDAR. *Forestry: An International Journal of Forest Research* 92:311–323. <https://doi.org/10.1093/forestry/cpz014>
- Zeide B (1980) Plot size optimization. *For Sci* 26:251–257. <https://doi.org/10.1093/forests/26.2.251>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.