https://doi.org/10.17221/141/2019-JFS

Retrieval of among-stand variances from one observation per stand

Steen Magnussen[1], Johannes Breidenbach[2]*

[1]Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria BC, Canada
[2]Norwegian Institute of Bioeconomy Research, Ås, Norway

**Corresponding author: steen.magnussen@canada.ca*

Citation: Magnussen S., Breidenbach J. (2020): Retrieval of among-stand variances from one observation per stand. J. For. Sci., 66: 133–149.

Abstract: Forest inventories provide predictions of stand means on a routine basis from models with auxiliary variables from remote sensing as predictors and response variables from field data. Many forest inventory sampling designs do not afford a direct estimation of the among-stand variance. As consequence, the confidence interval for a model-based prediction of a stand mean is typically too narrow. We propose a new method to compute (from empirical regression residuals) an among-stand variance under sample designs that stratify sample selections by an auxiliary variable, but otherwise do not allow a direct estimation of this variance. We test the method in simulated sampling from a complex artificial population with an age class structure. Two sampling designs are used (one-per-stratum, and quasi systematic), neither recognize stands. Among-stand estimates of variance obtained with the proposed method underestimated the actual variance by 30-50%, yet 95% confidence intervals for a stand mean achieved a coverage that was either slightly better or at par with the coverage achieved with empirical linear best unbiased estimates obtained under less efficient two-stage designs.

Keywords: coverage; heteroscedasticity; one-per-stratum; quasi systematic; stand mean; two-stage sampling

The availability of wall-to-wall coverage of remotely sensed data (**X** viz. auxiliary variables) correlated with one or more study variables of interest (Y) offers opportunities to improve the sampling designs of forest inventories (Duane et al. 2010; Tomppo et al. 2014; Fattorini 2015; Magnussen et al. 2017; Breidenbach et al. 2018). Sample selection protocols that ensure a uniform sampling density in the space spaned by **X** are becoming increasingly popular (Katila, Tomppo 2002; Grafström, Ringvall 2013; Grafström et al. 2014). Probability sampling designs ensuring a uniform dispersion of sample locations across a spatial population via a tessellation scheme as advocated by, for examples, Stevens and Olsen (2004) and Cordy (1993) are making inroads in natural resource surveys. Moreover, auxiliary variables provide opportunities for post-stratification (Dahlke et al. 2013; Ene et al. 2018; Räty et al. 2018).

Current forest enterprise inventory designs exploiting remotely sensed auxiliary variables appear to provide population level estimates with acceptable levels of precision (Von Lüpke et al. 2012; Corona et al. 2014; Melville et al. 2015; Mauro et al. 2016; Saarela et al. 2016). Yet a sampling design, tailored to provide accurate population level estimates, may not provide stand-level estimates with a precision required by management (Magnussen et al. 2016). Forest inventories typically have fewer sample plots than there are stands in the surveyed forest. Hence, most stands will have no information from field plots, and inference about stand means and their variances derives exclusively from

https://doi.org/10.17221/141/2019-JFS

the auxiliary variables and an assumed model fitted to the sample data. Unless the variance from stand effects is quantified, a model based variance estimator for a stand mean may woefully underestimate the actual variance (Breidenbach et al. 2016; Magnussen, Breidenbach 2017).

Stand-effects should be expected (Mäkelä, Pekkarinen 2004; Eerikäinen 2009; Nothdurft et al. 2009; Junttila et al. 2010; Goerndt et al. 2011; Mauro et al. 2016). Generalized mixed linear, semi-parametric, or non-parametric models with stands effects included as a random intercept, or as random interactions with one or more of the auxiliary variables in the model, have been used to predict stand effects and the variance arising from these effects (Jiang, Lahiri 2006; Bolker et al. 2009). Suitable sampling designs for these modelling approaches requires at least two sample plots in a minimum number of stands. A study by Maas and Hox (2005) suggests that 100 stands may be the minimum to achieve a correct coverage of nominal 95% confidence intervals for a stand mean.

For a given sample size, a sampling design that has at least two plots in a number of stands, will be less efficient in terms of the uncertainty in an estimate of a population total (mean) than a design without this constraint. We know this from stratified sampling where one sample per stratum is the most efficient design (Cochran 1977)(Self, Mauritsen 1988; Liu, Liang, 1997; Maas, Hox 2005; Snijders 2005).

Earlier studies have proposed methods for obtaining an estimate of the among-stand variance from sample data collected under a design that does not allow a direct estimation. In these cases the among-stand variance is estimated from predictions of the study variable (Magnussen 2018), or statistics such as the intra-cluster correlation coefficient whereby stands are treated as clusters (Magnussen 2016; Magnussen et al. 2016; Magnussen, Breidenbach 2017). In this study, we propose a new method for obtaining an estimate (from regression residuals) of the variance arising from stand-effects. We illustrate the method with two efficient sampling designs that stratify sample locations by an auxiliary variable, but do not otherwise allow a direct estimation of the among-stand variance.The first is the one-per-stratum (*OPS*) design proposed by Breidt et al. (2016), and the second is the quasi systematic sampling design (*QSYST*) proposed by Wilhelm et al. (2017). In the proposed new method, an estimate of the among-stand variance is obtained from a modified analysis of variance using regression residuals clustered on the basis of the auxiliary variable(s) used to allocate sample locations. Our illustration of the method is based on simulated sampling under the two designs from a a large artificial population with 1,000 stands in each of seven age classes, one study variable, and two auxiliary variables. A case study with an actual population was desired, but we have not been able to obtain data from a forest inventory with a study variable known for all live trees in a large number of stands.

Population and stand level results with *OPS* and *QSYST*, and the proposed new method for estimating the among-stand-variance, are compared to results with three variants of a two-stage sampling design with a direct (here: empirical best linear unbiased or EBLUP) estimation of the among-stand variance (Pinheiro, Bates 2000).

MATERIAL AND METHODS

Artificial population. Our artificial population is cast as a mid-sized forest estate of black spruce, the most common tree species in Canada. We assume a stand-oriented approach to forest management. That is reliable estimates for a population and stand means of key forest resources are required along with statistics of the uncertainty associated in these estimates, and their 95% confidence intervals. The artificial population should be large enough to allow simulated sampling with realistic sampling intensities for forest management inventories. Moreover, the artificial population should have an age-class structure and, for our purpose of demonstration, also include a practically important among-stand variance within each age class.

The artificial population (forest) covers an area of 10 100 ha, and has 17.9 million trees stratified to seven approximately equal area 20-year age classes and 1 000 stands in each age class. The stand sizes were chosen to emulate private forest holdings in eastern Canada and in much of Europe. That is, an average size of 1.4 ha and a standard deviation of 0.6 ha.

To enable simulated sampling under some chosen design, each stand is tessellated into a number of fixed area units with an area equal to the area of a field sample plot (here 200 m^2).

https://doi.org/10.17221/141/2019-JFS

The study variable (Y) is stem wood volume density (VOL $m^3 \cdot ha^{-1}$), and the two remotely sensed auxiliaries (details to follow) are cast as the canopy height (CHT m) of first-return pulses from an airborne laser scanner, and the within-plot standard deviation of the first-returns ($sCHT$ m). The two auxiliaries are often found to have good predictive powers of stem volume density (Magnussen et al. 2015). The limitation to two auxiliaries is argued on grounds of a parsimoneous model.

The trend over age in mean stem diameter (*mDBH*), expected height (*EHT*), and basal area density (BA $m^2 \cdot ha^{-1}$) follows Plonski's yield table (Plonski, 1960) for black spruce (*Picea mariana* (Mill.) BSP) for a medium productivity level site index (SI) of 18 m at age 100.

Next we generated 1 000 stand templates for each of the seven age classes (acl = 1, …, 7). Each template was populated in four steps:

(1) An age was assigned to a template in age class *acl* by a random draw from a discrete uniform distribution with lower and upper limits of 20 (*acl*-1) and 20 *acl*.

(2) A SI was assigned by a random draw on the interval 17.1–18.9 m using a generalized beta distribution (Fazar 1959; Li et al. 2002) with a mean of 18 and a standard deviation of 0.49.

(3) A mean *DBH* (*mDBH* cm) and a basal area (BA $m^2 \cdot ha^{-1}$) were determined from functions of SI and age (Payandeh 1991) after slight modifications to ensure a monotone trend over the age interval 15–154 years.

(4) An expected stem density (*ESTEMS* $stems \cdot m^{-2}$) was obtained from *BA* and *mDBH*, and multiplied by a correction factor to ensure a non-increasing trend over age. The 7 000 stand templates were then converted to virtual stands with randomly chosen areas (*SA*) restricted to a multiple of 200 m^2. All stand areas are between 0.4 ha and 2.6 ha with a mean of 1.4 ha. Thus the average number of 200 m^2 units in a stand is 70 (min = 20, max = 130). The number of stems (*STEMS*) in a stand was decided by a random draw from a Poisson distribution with mean $SA \times ESTEMS$ and truncated at $0.75 \times SA \times ESTEMS$ and $1.33 \times SA \times ESTEMS$. Our choice of truncation points controls the coefficient of variation in the number of stems in a stand. Specifically, the approximate coefficient of variation in stem numbers in the seven age-classes is: 0.08, 0.10, 0.11, 0.12, 0.13, 0.13, and 0.13.

To generate a within stand variation in stem density, the *STEMS* in a stand with an area composed of $n_p \times 200$ m^2 units were distributed to the units according to a set of random proportions. The n_p random proportions were obtained from a Dirichlet distribution with all n_p parameters equal to one (Darroch, Ratcliff 1971). Thus the expected proportion of STEMS in a unit is n_p^{-1} with a variance of $(n_p - 1)(1 + n_p)^{-1} \times n_p^{-2}$ which declines in n_p. To wit, the standard deviation with n_p = 30 is 0.032, 0.014 with n_p = 70, and 0.008 with n_p = 130. The number of stems in the n_p units was hereafter computed as the rounded value of the product of *STEMS* and the randomly drawn proportions.

To populate the 17.9 million stems with values of *DBH*, tree height (*HT* m), and total stem volume (*VOL* m^3), we first generated stand specific models for the distribution of *DBH*. We chose a three parameter Weibull distribution. A model often used for this purpose (for examples, Bailey, Dell 1973; Magnussen 1986; Cao 2004). The stand specific parameters of this distribution was obtained by the method of moments from the known value of *mDBH*, assuming a coefficient of variation of 33%, and equating the lower limit to the 1/*STEMS* quantile of student's t-distribution with mean *mDBH*, standard deviation of 0.33 × *mDBH*, and *STEMS* degrees of freedom. The maximum *DBH* value in a stand was equated to the 99-percentile of the assumed Weibull distribution. The fitted Weibull distribution was then used to draw individual stem values of *DBH* and *HT* from a data-base of 10 000 such pairs (see next) with probability proportional to the *DBH* values predicated by the Weibull distribution. Volume values were obtained from a volume equation (see next) and the drawn values of *DBH* and *HT*. Additional stand effects (beyond those arising from *SI*, age, and *STEMS*) were added by multiplying each *HT*-value in a stand by $(1+\Delta\ HT)$ with Δ representing a random draw from a uniform distribution on the interval [–0.05, 0.05], and by multiplying each *VOL* value by $(1+\delta\ VOL)$ with δ representing a random draw from a uniform distribution on the interval [–0.07, 0.07].

The data-base with 10 000 paired values of *DBH* and *HT* was constructed from 785 felled black spruce trees collected across its range in Canada (Macleod 1978); the largest data set available for this species. The tree data include a measured height (*HT*), a stem diameter at a reference level of 1.3 m above ground (*DBH* cm), and a total stem volume

https://doi.org/10.17221/141/2019-JFS

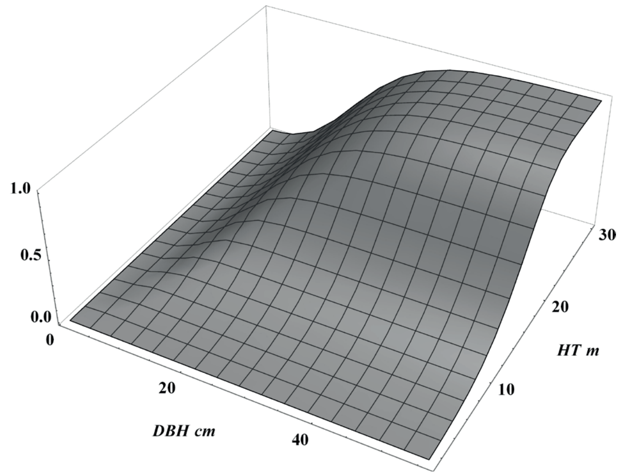

Figure 1. The cumulative bivariate distribution function of DBH and HT

(*VOL* m^3) computed via Smalian's formula applied to 2-m stem sections (Spurr 1952). The average tree size was 18.9 cm for *DBH* with a standard deviation (*sdev*) of 8.6 cm, and a range from 2 to 54 cm. Tree heights varied from 1.6 m to 25.5 m with an average of 14.9 m, and a *sdev* of 4.4 m. Total stem volume varied from 0.0017 m^3 to 2.397 m^3 with a mean of 0.272 m^3 and a *sdev* of 0.282 m^3.

The 785 paired values of *DBH* and *HT* were bulked to 10 000 pairs by random draws from a bivariate Copula distribution of *DBH* and *HT* fitted to data by methods of maximum likelihood (Fischer 2010). The best fit marginal distribution of *DBH* was a Johnson's SL distribution (Rennolls, Wang 2005) with parameters shape $\gamma = 28.7382$, shape $\delta = 3.77002$, location $\mu = -13.2$, and scale $\sigma = 63{,}318.7$. The best fit marginal distribution for *HT* was a truncated exponential gamma-distribution (Dubey 1970) with parameters shape $\kappa = 1.34274$, scale $\phi = 4.71605$, location $\mu = 15.2022$, and a left truncation at 1.3 m. A Clayton kernel with a parameter value of 0.31 was used to tie the two marginal distributions together to a bivariate distribution. The fitted bivariate cumulative distribution function of *DBH* and *HT* is in Figure 1. For each drawn pair of *DBH* and *HT*, the value of *VOL* came from application of Evert's volume equation (Evert 1983) that was fitted using the data from the 785 felled trees.

Simulated auxiliary variables. Two auxiliary variables are used in unit-level (i.e. area based) model predictions of *mVOL* for every unit in the artificial population. The two auxiliaries are intended to emulate two LiDAR metrics of first-return echoes obtained wall-to-wall from an airborne laser scanner (Sexton et al. 2009; Næsset 2014). The first auxiliary is cast as the mean canopy height (*mCHT*) (Sexton et al. 2009). A *CHT* value was generated for each stem in the artificial population. Specifically. $CHT = HT - 1.5 + rHT$ where *rHT* is a uniform distributed random variable on the interval –0.5 m to 0.5 m. All generated values of *CHT* were subsequently censored to the interval [1.3, 26.0] m and averaged to *mCHT* for each 200 m^2 unit. The second auxiliary *sCHT* is the standard deviation in *CHT* in a 200 m^2 unit.

Properties of the artificial population. Table 1 provides summary statistics by age class and for age class one to seven combined.

The among-stand variances in Table 1, expressed as a coefficient of variation (CV, i.e. the stratum among-stand standard deviation divided by the stratum mean), varied from 12% in age class one to 16% in age class seven, and was 17% over all classes. The among-stand variance in *mCHT* was greatest in age class one with a CV of 36%. In age class two to seven, the CV dropped from 10% to 4%. In *sCHT* the CV was 32% for age class one and approximately 10% in age class two to seven.

The coefficient of variation in the age class specific unit-level distributions of *mVOL* $m^3 \cdot ha^{-1}$ was highest in age class one (58%), and it then tapered off to approximately 30% in the older classes. For *mCHT* the CV was 38% in age class one, 11% in age class two, and then dropping to 7% and 5% in age classes three to seven. Newton and Amponsah (2007) noted a comparable decline.

Age class specific distributions of *mVOL* $m^3 \cdot ha^{-1}$ were right-skewed. The skewness coefficient dropped from 0.76 to 0.52 over the first three age classes, and then hovered at this level in the four older classes. Skewness was unimportant in *mCHT*, and between 0.15 and 0.23 in *sCHT*.

Sampling designs. We test the proposed new method for obtaining an estimate of the among-stand variance in two efficient sampling designs that stratify sampling locations by an auxiliary variable (*mCHT*), but do not otherwise allow a direct estimation of this variance. Results with the new method are compared to results from two-stage sampling designs (*2ST*) that allow a direct computation of the among-stand variance. All sampling is by age-class, and results are provided for each age class, and for the population. Specifically, the sampling designs for testing the new method are stratified one-per stratum (*OPS*) and stratified quasi systematic sampling (*QSYST*), neither has been explored for forest inventories, and in the opinion of the authors, they deserve attention. The

https://doi.org/10.17221/141/2019-JFS

Table 1. Summary statistics from the artificial black spruce forest

	Age class							
	1	2	3	4	5	6	7	1–7
Area (ha)	1 452	1 429	1 434	1 483	1 431	1 463	1 429	10 100
No. of stands	1 000	1 000	1 000	1 000	1 000	1 000	1 000	7 000
Avg. stand area	1.5	1.4	1.4	1.5	1.4	1.5	1.4	1.4
mVOL ($m^3 \cdot ha^{-1}$)								
Mean	57	172	246	292	326	354	377	261
sdev	33	52	71	87	100	109	118	136
Skewness	0.76	0.66	0.52	0.49	0.53	0.51	0.55	0.31
Among-stand variance	47	396	1 038	1 648	2 299	2 870	3 482	1 889
mCHT (m)								
Mean	5.3	10.6	12.9	14.0	14.6	14.8	15.1	12.5
sdev	2.0	1.2	0.9	0.9	0.8	0.8	0.8	3.5
Skewness	0.12	–0.02	0.05	0.01	–0.01	–0.03	–0.03	–1.4
Among-stand variance	3.6	1.2	0.5	0.3	0.3	0.3	0.3	11.6
sCHT (m)								
Mean	2.2	3.4	3.4	3.3	3.2	3.1	3.1	3.1
sdev	0.7	0.5	0.5	0.6	0.6	0.5	0.5	0.7
Skewness	0.21	0.23	0.15	0.15	0.19	0.18	0.22	0.43
Among-stand variance	0.5	0.1	0.1	0.1	0.1	0.1	0.1	0.3
Correlation (*mVOL*,*mCHT*)	0.91	0.63	0.43	0.36	0.34	0.32	0.31	0.78

mVOL – model wood volume density, *mCHT* – *model* auxiliary variable, *sCHT* – standard deviation of the first-returns

sampling designs are detailed in the next section. Three population level sampling intensities (MIN, MED, and MAX) assumed compatible with forest enterprise inventories are executed. We have n = 542 units for MIN viz. one unit per 18.6 ha forest, n = 1 076 for MED viz. one per 9.4 ha, and n = 2 146 for MAX viz. one per 4.7 ha. Age class specific sample sizes are in Table 2. They were determined by Neyman allocation (Snedecor, Cochran 1971) with proportionality to the product of stratum area and the known standard deviation of *mCHT* in each age-class.

Table 2. Two-stage sampling design (2ST) sample sizes (units) in *2ST* variants MIN, MED, MAX, and number of selected stands in MIN, MED, and MAX variants *2ST100*, *2ST60*, and *2ST30*

	Age class							
	Sample Size (units)							
2ST								
MIN	30	44	68	86	96	104	114	542
MED	60	86	136	170	190	208	226	1076
MAX	120	170	270	340	380	416	450	2 146
	Number of stands with two sample units							
2ST100								
MIN	15	22	34	43	48	52	57	271
MED	30	43	68	85	95	104	113	538
MAX	60	85	135	170	190	208	225	1 073
2ST60								
MIN	10	14	18	24	30	34	38	168
MED	20	30	44	50	56	62	68	330
MAX	36	50	80	102	114	124	134	640
2ST30								
MIN	5	7	9	12	15	17	19	84
MED	10	15	22	25	28	31	34	165
MAX	18	25	40	51	57	62	67	320

Stratified one-per-stratum sampling

In the one-per-stratum (*OPS*) sampling design (Cochran 1977), age-class specific sample sizes (n_h) are equal to those of the *2ST* variants MIN, MED, and MAX in Table 2. To implement *OPS*, each age-class is stratified into n_h strata by the remotely sensed auxiliary variable with the strongest correlation to *mVOL* i.e. *mCHT*. We chose an efficient stratification by the $CUM\sqrt{f}\,(x)$ rule of Dalenius and Hodges (1959), and detailed in Cochran (1977). Hence, within an age class, the *OPS* sample distribution of *mCHT* values is balanced against the distribution of *mCHT* in an age class.

Quasi systematic sampling

In *QSYST* the age class specific sample sizes (n_h) are equal to the *2ST* variants MIN, MED, and MAX in Table 2. The objective is to obtain, for each age

https://doi.org/10.17221/141/2019-JFS

class, a uniform sample distribution of the auxiliary variable with the strongest correlation to *mVOL*, i.e. *mCHT.* We followed the sampling protocol in Wilhelm et al. (2017). Selections are initially carried out with an ancillery variable *z* uniformly distributed on the interval [0; 1]. The within age class sample selection was hereafter obtained as per Equation 3 in Wilhelm et al. (2017) for a binomial renewal process and a tuning parameter of $r = 2.5$. The conversion from selected *z*-values to *mCHT*, was implemented via a nearest neighbour search in paired values of (*z*(*mCHT*), *mCHT*) where *z*(*mCHT*) is the transform of *mCHT* to the interval [0; 1]. Anderson-Darling tests (Anderson, Darling 1952) of a uniform distribution of drawn *z*-values on the interval [0, 1] confirmed the workings of the algorithm.

According to theory, a larger *r*-value would have been preferred, since sampling becomes strictly systematic as *r* goes to infinity. However, for the sample sizes in this study, we ran into numerical accuracy issues and excessive computation times when attempting to compute the joint sampling densities in (9) with *r* values greater than 10. Our choice of *r* is therefore a compromise between asymptotics, and the risk of selecting an age class unit more than once – which increases as *r* decreases. With $r = 2.5$ we encountered no duplicates in a total of 2 100 randomly drawn samples.

Two-stage sampling designs

In stage one of the two-stage benchmark designs, m_h stands are randomly selected (without replacement, *wor*) in age class *h* (*h* = 1, …, 7), and in stage two, either one or two units are selected at random (*wor*) from each of the m_h stands selected in stage one (Cochran 1977). We included three variants of two-stage sampling. In the variant called *2ST100*, two units are selected from each selected stand; in the variant called *2ST60*, two units are selected in approximately 60% of the m_h stands, and one unit is selected in each of the remaining sampled stands; in the third variant called *2ST30*, two units are selected in approximately 30% of the selected stands, and one unit from each of the remaining. The m_{h2} (m_{h1}) stands from which to select two (one) units were selected at random (*wor*) from the set of m_h stands selected in stage one. Table 2 provides details.

In the simulated sampling (and in all estimators), sample data consist of unit (plot) means *mVOL* m^3·ha^{-1} of *VOL* and *mCHT*, and the within-plot standard deviation *sCHT* m of *CHT*.

Inference. We pursue design-based inference about the population mean, age class means, and model-based inference about stand means of *mVOL*. In each case, we seek to quantify the uncertainty (standard error) in an estimated mean, and we compute a nominal 95% confidence interval for the mean. The proposed new method for estimation of the among-stand-variance is only relevant for the stand-level inference. However, since *OPS*, *QSYST*, and even *2ST* are not common sampling designs in enterprise forest inventories, we do include population and age class results in our exposé. Readers interested only in the stand-level inference may skip details of estimators used for population and age class level inference.

All levels of inference employ a linear assisting model stating an assumed (not necessarily true) relationship between *mVOL* and the two remotely sensed auxiliaries *mCHT* and *sCHT*. To wit (Eq. 1):

$$y_{hi} = \beta_{1h} mCHT_{hi} + \beta_{2h} sCHT_{hi} + \varepsilon_{hi}, \quad i = 1,\ldots,N_h, h = 1,\ldots,7 \tag{1}$$

where y_{hi} the value of *mVOL* m^3·ha^{-1} in the i^{th} unit in age class *h*, β_{1h} and β_{2h} are regression coefficients to be estimated from sample data, and ε_{hi} is an error assumed normally distributed with a mean of zero and an assumed variance $\sigma_{hi}^2 = exp(\alpha_0 + \alpha_1\, mCHT_{hi})$. Note, the error term in (1) includes a stand effect. In estimators for the two-stage designs, the error term is decomposed to a term linked to the stand effect, and a term linked to the within stand error (details are in the subsections for the *2ST* designs). With our proposed new method we achieve, as detailed below, a decomposition via a modified analysis of variance of the error term in Eq. (1).

Estimators for population and age class inference. Estimators of age class and population means of *y* and their uncertainties require predictions of and empirical residuals for all units in a sample ($h = 1,\ldots,7$, $i = 1,\ldots, n_h$). Predictions came from (Eq. 2):

$$\hat{y}_{hi} = \hat{\beta}_{1h} mCHT_{hi} + \hat{\beta}_{2h} sCHT_{hi}, i = 1,\ldots,N_h, h = 1,\ldots,7 \tag{2}$$

where $\hat{\beta}_{1h}, \hat{\beta}_{2h}$ are design-consistent regression coefficients estimated via weighted least squares regression (Särndal et al. 1992). Let w_{hk} denote the weight assigned to the k^{th} unit in age class *h*. Under a *2ST* design w_{hk} is a sample inclusion probability, under *OPS* a $CUM\sqrt{f}(x)$ stratum weight, and under *QSYST* a sampling density (Wilhelm et al. 2017). Specifically (Eq. 3):

https://doi.org/10.17221/141/2019-JFS

$$\left(\hat{\beta}_{1h},\hat{\beta}_{2h}\right)=\left(\frac{\sum_{k=1}^{n_h}\langle mCHT_{hk},sCHT_{hk}\rangle\langle mCHT_{hk},sCHT_{hk}\rangle^t}{w_{hk}\,\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)}\right)^{-1}$$

$$\sum_{k=1}^{n_h}\frac{\langle mCHT_{hk},sCHT_{hk}\rangle y_{hk}}{w_{hk}\,\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)} \tag{3}$$

where the superscript 't' denotes a transpose of a vector, and $\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)$ is an estimator of the residual variance in ε_{hk} expressed as an exponential function of $mCHT_{hk}$. The variance function $\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)=\exp\left(\hat{\theta}_{h0}+\hat{\theta}_{h1}mCHT_{hk}\right)$ was estimated by a four-step method (Harvey 1976). In step one, ordinary least squares (OLS) estimates of β_{1h}, β_{2h} are obtained; in step two, OLS estimates of $\hat{\theta}_{h0},\hat{\theta}_{h1}$ are obtained via a linear regression with the log of squared residuals from step one as the dependent variable; in step 3 a bias correction is added to $\hat{\theta}_{h0}$; and in step four, generalized least squares estimates of β_{1h}, β_{2h} are obtained with the fitted variance function.

The regression estimator of an age class mean of y was hereafter (Särndal et al. 1992), Equation 4:

$$\bar{y}_h\left(DES\right)=\sum_{hk=1}^{N_h}\hat{y}_{hk}+\sum_{s\ni k}\frac{\hat{\varepsilon}_{hk}}{w_{hk}},DES=2ST,OPS,QSYST \tag{4}$$

Estimates of the population mean was obtained as the weighted sums of age class means (Cochran 1977).

Under the *OPS* design, the estimator of the variance in the mean of y in age class h is (Breidt et al. 2016), Equation 5:

$$\hat{V}\langle\bar{y}_h\left(OPS\right)\rangle=\frac{1}{c_{dh}}\sum_{k=1}^{n_h}W^2_{hk}\left(y_{hk}-\sum_{hl=1}^{n_h}d_{hk}\left(hl\right)y_{hl}\right)^2,h=1,\ldots,7 \tag{5}$$

where W_{hk} is the area proportion of age class h units in the k^{th} $CUM\sqrt{f}\,(x)$ stratum (Breidt et al. 2016, Eq. (6)), d_{hk} (hl) is a kernel weight for the sample data pair of the k^{th} and l^{th} unit in the k^{th} $CUM\sqrt{f}\,(x)$ stratum in age class h, and is a normalizing constant. We chose a triangular kernel ($K(x) = (1 - |x|)$ if $|x| \leq 1$ and 0 elsewhere), Izenman 1991 based on its root mean squared error in a leave-one-out cross-validation. The weights d_{hk} (hl) in (Eq. 5) comes from Equation 6:

$$d_{hk}\left(hl\right)=\frac{K\left(\frac{mCHT_{hk}-mCHT_{hl}}{\hat{\omega}_h}\right)}{\sum_{l=1}^{n_h}K\left(\frac{mCHT_{hk}-mCHT_{hl}}{\hat{\omega}_h}\right)} \tag{6}$$

And the bias-correction c_{dh} from Equation 7:

$$c_{dh}=n_h^{-1}\sum_{k=1}^{n_h}\left(1-2d_{hk}\left(hk\right)+\sum_{l=1}^{n_h}d^2_{hl}\left(hk\right)\right) \tag{7}$$

The kernel bandwidth $\hat{\omega}_h$ in (Eq. 6) minimized the mean squared error of the kernel-weighted sample mean of *mCHT* in age class h (h = 1, …, 7).

The approximate variance of $\bar{y}_h(QSYST)$ is (Wilhelm et al. 2017) (Equation 8):

$$\hat{V}(\bar{y}_h\left(QSYST\right))=\sum_{k=1}^{n_h}\sum_{l=1}^{n_h}\frac{y_{hl}}{\tilde{\pi}_{hk}}\frac{y_{hk}}{\tilde{\pi}_{hl}}\frac{\left(\tilde{\pi}_{hk,hl}-\tilde{\pi}_{hk}\tilde{\pi}_{hl}\right)}{\tilde{\pi}_{hk,hl}} \tag{8}$$

with $\tilde{\pi}_{hk}$ denoting a sampling density (Cordy 1993) whereby $\tilde{\pi}_{hk}=\tilde{\pi}_{hl}=n_h$, and the joint sampling density $\tilde{\pi}_{hk}$, $\tilde{\pi}_{hl}$ is a function of the absolute difference in the draws of the scaled random variable z (see sampling designs). For hk = hl we get $\tilde{\pi}_{hk,hl}=1-n_h^2$ and for $hk \neq hl$ (Equation 9):

$$\tilde{\pi}_{hk,hl}=n_h\sum_{m=1}^{n_h}\frac{\Gamma\left(n_h r\right)}{\Gamma\left(mr\right)\Gamma\left(n_h-m\right)r}\left|z_{hk}-z_{hl}\right|^{mr-1}\left(1-\left|z_{hk}-z_{hl}\right|\right)^{(n_h-m)r-1} \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function. As expected (Wilhelm et al. 2017), we have $\tilde{\pi}_{hk,hl}\cong n_h\,(n_h-1)$ for nearly all pairs of z_{hk} and z_{hl}.

The variance of an age class mean obtained under a two-stage sampling design was estimated as per 4.3.14 in Särndal et al. (1992) and is not detailed here.

The variance of a population mean was in all cases obtained as the weighted average of the age class variances (Cochran 1977).

Stand-level inference. With the estimated age class specific regression coefficients, a regression-synthetic estimate of the mean of $mVOL$ m^3·ha^{-1} in a stand in age class $h(\bar{y}_{hst})$ was obtained for all stands. The estimated stand mean is simply the mean of unit-level predictions of y computed for all N_{hst} units in a stand. The variance estimator of $\bar{y}_{hst}$ is in Equation 10:

$$\hat{V}\left(\bar{y}_{hst}\right)=\frac{1}{N_{hst}}\sum_{st\ni hk}\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)mCHT_{hk},\cdot$$
$$sCHT_{hk}\left(\boldsymbol{XX^t}\right)^{-1}mCHT_{hk},sCHT_{hk}{}^t+ \tag{10}$$
$$+\hat{\sigma}^2_{hst}+\frac{\left(\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)-\hat{\sigma}^2_{hst}\right)}{N_{hst}},h=1,\ldots,7,st=1,\ldots,1\,000$$

where $(XX^t)^{-1}$ is the inverse of the 2 × 2 matrix of cross product sums of $mCHT_{hk}$ and $sCHT_{hk}$ in the age class from which the regression coefficients were derived, $\hat{\sigma}^2_{hst}$ (see details in next section) is an estimate of the among-stand variance in age class h, and $\hat{\sigma}^2_{hk}\left(mCHT_{hk}\right)$ is the predicted residual variance applicable to unit k with a mean canopy height $mCHT_{hk}$. To illlustrate the importance of the among-stand variance for the coverage of a nominal 95% confidence interval, the variance in (Eq. 10) was computed twice,

https://doi.org/10.17221/141/2019-JFS

first with $\hat{\sigma}_{hst}^2$ = 0 and then with the obtained estimate $\hat{\sigma}_{hst}^2$.

A nominal normal 95% confidence interval (CI95) for the true stand mean was obtained from $\bar{y}_{hs}$and $\hat{V}(\bar{y}_{hst})$ by standard techniques (Casella, Berger 2002). As for the variance in Equation 11, two confidence intervals were computed.

With the *2ST* designs, the among-stand variance $\hat{\sigma}_{hst}^2$ was computed via restricted maximum likelihood of a mixed linear model with stand as random effects and *mCHT* and *sCHT* as covariates (Pinheiro, Bates 2000). For *OPS* and *QSYST*, this method is not feasible since it is rare to have more than one sampled unit in a stand. For these designs the among-stand variance is obtained with the proposed new method.

The tenent behind our new method is is that stands with a similar *mCHT* value also have similar (random) stand effects. The same tenet underpins the traditional estimation of variance in *OPS* with collapsed strata (Snedecor, Cochran 1971). Therefore, if we cluster the empirical residuals (cf Eq. 1) based on their *mCHT* values, we should ideally expect, that the stand effects would more similar within a cluster than if clustered by a random assignments. A one-way analysis of variance (ANOVA) with the clustered empirical residuals would, ideally, provide us with a reasonable proxy to the desired among-stand variance. The key question as to how many clusters (Fraley, Raftery 1998) can not be answered by theory. Preliminary investigations suggested that more than one cluster size would be needed to obtain robust estimates, and also that data-driven modifications to the analysis of variance would be needed. Details are next.

Specifically, we propose the following estimator of the among-stand variance

$$\hat{\sigma}_{hst}^2(DES) = \left(n_{clu}^{\alpha 1}\mathrm{MSE}_{clu} - \mathrm{MSE}_{w}\right)/n_{clu}^{\alpha 2}\Big), \quad DES = \{OPS, QSYST\} \qquad (11)$$

where n_{clu} is the number of empirical residuals in a cluster, MSE_w is the ANOVA within-cluster mean squared residual error, and MSE_{clu} is the ANOVA among-cluster mean squared residual error. The parameters α1 and α2 are to be obtained by data-driven simulations (see next). Here we use α1 = 1 and α2 = 2.4 (see next for details). The estimator in (11) was derived for cluster sizes (n_{clu}) of 2, 3, 4, and 5 and then averaged to a final estimate. We found that the average of the four estimates of an among-stand variance bestowed robustness to the estimated variance.

In applications, the parameter values for α1 and α2 must be obtained from simulations with standard bivariate Gaussian variables, whereby the second is a sum of two zero-mean independent Gaussian terms emulating random zero-mean stand effects with a variance σ^2_{stand}, and a residual errors with a mean of zero and a variance of $1 - \sigma^2_{\text{stand}}$. The bivariate correlation was, in our simulations, fixed at $r = 0.50$, but values between 0.1 and 0.9 will also work. Specifically, for $\sigma^2_{\text{stand}} = 0.05$, 0.10, …, 0.30, and sample sizes = 50, 100, …, 500, we split the sorted (by the first Gaussian variable) sample into clusters of size 2, 3, 4, or 5 and obtained MSE_{clu} and MSE_w by standard ANOVA techiques (Snedecor, Cochran 1971) for each combination of σ^2_{stand}, cluster size, and sample size. This process was repeated 100 times. With σ^2_{stand} known, we estimated α1 and α2 by minimizing the sum of squared differences over the 24, paired values of the variances on the right and left hand side of (Eq. 11). The pseudo R^2 (Mbachu et al. 2012) between the actual among-stand variance and the estimated variance was 0.48 in our 24 000 simulated cases.

We used Levene's test of equal variances (Levene, 1960) to compare estimates of the among-stand variance and the distribution of *p*-values from these test were compared to a uniform distribution on the closed interval [0; 1] as expected under the null hypothesis (Anderson, Darling 1952).

Performance criteria. The sampling designs are evaluated in terms of bias in an estimate of a mean, empirical standard errors (ESE) of a mean, the correspondence between an empirical and an analytical standard error (ASE), and the coverage of nominal 95% confidence intervals.

Bias was computed as the difference between an estimated mean and the actual value in percent of the actual mean. This bias is labelled BIAS%. An ESE is computed as the standard deviation of the 100 estimates of a mean for a given combination of design × sample size divided by √100. The efficiency of a design is measured by its ESE. The lower the ESE, the higher the efficiency. ASE is the squared root of the average estimated variance in an estimate of a mean.

RESULTS

Age class and population inference

The main objective of this section is to illustrate the efficency of the *OPS* and *QSYST* sampling de-

https://doi.org/10.17221/141/2019-JFS

signs. Readers mainly interested in our new method for estimating an among stand variance from designs that do not allow a direct estimation may skip the remainder of this section.

The relative bias in an estimates of an age class mean of *mVOL* m^3·ha^{-1} is summarized in Table 3. In age class three to seven, the BIAS% was similar across the five designs and sample sizes (MIN, MED, MAX). In age class one and two, BIAS% was more variable. The higher relative variability and skewness in *mVOL* m^3·ha^{-1} in the first two age classes points to the challenge of obtaining accurate estimates with our sample sizes in presence of skewness. We learned that approximately 200 replications of a sampling design would be needed to increase the statistical power of the *t*-tests to declare a bias of 1% (or greater) as statistically significant at the 5% level.

Empirical standard errors in the estimates of the age class means of *mVOL* m^3·ha^{-1} (obtained from the 100 replications of a design × sample size combination) obtained with *OPS* and *QSYST* were approximately 40% lower than with the best two-stage variant *2ST30* (Table 4). The two-stage variant *2ST100* was, as expected, the least efficient. Over all settings, the empirical standard errors with *2ST100* were 13% greater than with *2ST30*. The higher efficiency of *OPS* and *QSYST* is upheld across age classes and sample sizes (MIN, MED, MIN). Differences between the *OPS* and *QSYST* designs in empirical standard errors did not indicate that one was more efficient than the other.

The mean of the analytical estimates of the standard error under the *OPS* and *QSYST* designs matched the empirical counterparts to within a few (2) percentage points, and the correlation between the two sets of estimates was strong (0.999). For the two-stage design *2ST30*, the correlation was somewhat weaker (0.91) but there was no apparent overall difference between the two error estimates. For *2ST60* and *2ST100*, the correlation was even lower (0.85), with the mean analytical error trailing the empirical error by 5% to 12% across all age classes and sample size levels.

Coverage rates for *OPS* and *QSYST* were between 89% and 98%. We failed to reject the null hypothesis of a coverage of 95% for these two designs. The two-stage design variant *2ST30* achieved the best MA coverage with an average matching the nominal 95% level. With the *2ST60* variant the average coverage was 93% but only 90% with *2ST100*. The minimum coverage (0.87) was with *2ST100* and a MIN sample size.

Stand-level inference

The assumed model in Eq. (1) is intended for population level inference. When used to obtain predictions of stand means, we should expect that the dif-

Table 3. Relative bias (BIAS%) in estimates of mean volume (m^3·ha^{-1}) by sampling design and age class

Design	Sample size	Age class							
		1	2	3	4	5	6	7	1–7
2ST100	MIN	–0.22	–0.31	–0.17	0.14	0.27	–0.30	–0.16	–0.08
	MED	**–0.91**	0.52	–0.10	–0.27	**0.45**	**–0.57**	0.23	–0.18
	MAX	**–0.85**	–0.22	–0.17	–0.20	–0.01	–0.14	0.20	–0.09
2ST60	MIN	–0.25	–0.13	–0.27	–0.15	0.01	–0.55	–0.09	–0.21
	MED	0.05	**–0.66**	**–0.58**	–0.24	0.12	–0.44	–0.30	–0.30
	MAX	**–0.98**	–0.32	0.30	0.05	0.01	0.13	0.26	0.05
2ST30	MIN	–0.18	–0.33	0.01	0.34	–0.02	0.10	–0.56	–0.19
	MED	–0.44	–0.20	–0.07	**–0.49**	0.02	0.02	–0.20	–0.16
	MAX	**–0.74**	0.11	–0.11	0.05	–0.03	0.04	–0.26	–0.07
OPS	MIN	0.01	–0.50	0.24	–0.52	0.29	–0.46	0.32	–0.27
	MED	**–0.66**	0.11	0.15	0.34	0.46	–0.11	–0.30	0.01
	MAX	**–0.80**	0.17	0.10	0.00	0.13	–0.23	–0.14	–0.05
QSYST	MIN	–0.54	–0.91	–0.5	0.23	0.56	–0.07	–0.08	–0.04
	MED	–0.77	–0.11	0.05	–0.41	–0.28	0.05	0.18	–0.08
	MAX	–0.87	–0.27	–0.34	0.07	–0.05	–0.17	0.12	–0.10

BIAS% – (estimated – actual)/actual × 100%, in bold – bias with a significant (5% level or less) departure from 0% in a t-test of the null hypothesis of a zero bias, *2ST100, 60, 30* – two-stage sampling designs, *OPS* – one-per-stratum, *QSYST* – quasi systematic sampling design

https://doi.org/10.17221/141/2019-JFS

Table 4. Analytical (ASE) and empirical (ESE) estimates of standard error in an estimate of an age class mean of *mVOL* m^3·ha^{-1} obtained under five designs and three levels of stratum sample sizes (MIN, MED, and MAX). Table entries are listed as ASE/ESE

Design	Sample size	Age class							
		1	2	3	4	5	6	7	1–7
2ST100	MIN	3.0/4.1	6.9/7.8	8.3/11.4	9.5/11.8	10.3/11.5	10.9/13.1	11.5/13.9	8.6/10.5
	MED	2.1/2.6	4.8/4.8	6.0/7.0	6.7/7.1	7.6/7.6	7.8/7.2	8.1/8.6	6.2/6.4
	MAX	1.5/1.7	3.5/4.1	4.3/4.9	4.8/5.5	5.3/6.3	5.5/5.6	5.8/5.9	4.4/4.0
2ST60	MIN	2.9/3.3	7.1/8.0	8.9/9.1	10.0/9.9	11.1/11.5	11.5/12.7	11.9/13.9	9.1/9.8
	MED	2.2/2.3	5.1/5.7	6.4/6.1	7.2/7.4	8.0/7.1	8.3/8.0	8.7/9.1	6.6/6.5
	MAX	1.6/1.9	3.7/4.1	4.6/4.6	5.0/5.4	5.6/6.3	5.8/5.9	6.0/6.3	4.6/4.9
2ST30	MIN	3.2/3.3	6.9/7.2	8.9/9.0	9.7/10.4	10.9/10.7	11.1/12.0	11.9/13.4	8.9/9.4
	MED	2.1/2.3	5.3/4.2	6.1/5.8	6.9/6.2	7.8/8.1	8.0/8.2	8.4/8.2	6.4/6.1
	MAX	1.6/1.7	3.6/3.4	4.4/4.1	5.0/4.7	5.5/4.9	5.7/5.4	6.0/6.0	4.5/4.3
OPS	MIN	2.9/2.7	6.6/6.5	8.2/7.8	9.2/8.3	10.3/9.9	10.9/12.1	11.1/11.0	8.4/8.3
	MED	2.0/1.8	4.7/4.1	5.7/5.2	6.6/5.8	7.2/8.0	7.5/7.7	7.8/8.0	5.9/5.8
	MAX	1.4/1.3	3.3/3.2	4.1/4.5	4.6/4.4	5.1/5.2	5.3/5.1	5.5/5.8	4.2/4.2
QSYST	MIN	2.9/2.7	6.4/6.4	7.9/7.0	8.7/8.4	9.6/9.7	10.4/10.3	10.8/11.4	8.1/8.0
	MED	2.0/2.0	4.6/4.7	5.4/5.8	6.2/6.2	7.0/6.7	7.1/6.8	7.6/7.3	5.7/5.6
	MAX	1.4/1.5	3.2/3.1	4.0/3.8	4.4/4.8	4.9/4.8	5.1/5.0	5.2/4.6	4.0/3.9

2ST100, 60, 30 – two-stage sampling designs, *OPS* – one-per-stratum, *QSYST* – quasi systematic sampling design

Table 5. Coverage of estimates of 95% confidence intervals for the actual mean *mVOL* m^3·ha^{-1} (coverage is in percent of 100 direct estimated intervals which include the actual mean)

Design	Sample size	Strata							
		1	2	3	4	5	6	7	1–7
2ST100	MIN	81	91	80	87	89	91	93	100
	MED	85	96	89	93	94	96	95	100
	MAX	91	88	91	87	89	96	94	100
2ST60	MIN	88	91	91	94	94	91	88	100
	MED	90	93	98	95	98	97	94	100
	MAX	89	90	95	94	92	96	93	100
2ST30	MIN	91	96	94	91	95	90	96	100
	MED	91	97	95	97	95	94	95	100
	MAX	93	97	95	97	94	96	96	100
OPS	MIN	96	94	98	98	93	97	95	100
	MED	96	96	94	97	92	96	96	100
	MAX	98	92	91	99	91	96	95	100
QSYST	MIN	89	91	97	95	94	93	91	100
	MED	94	94	89	94	94	98	93	100
	MAX	92	93	94	95	95	94	97	100

2ST100, 60, 30 – two-stage sampling designs, *OPS* – one-per-stratum, *QSYST* – quasi systematic sampling design

ference between an actual and a predicted mean (i.e. bias) can be important for some stands. Bias impacts coverage of nominal confidence intervals (Snedecor, Cochran 1971, Table 5). Here, an absolute bias of 10% or less had only a weak impact on coverage of a nominal 95% confidence interval, but with greater absolute bias, an increase of one percentage point incurred a loss in coverage between 0.002 and 0.03. There was no apparent trend in this loss attributable to age class, sample size, or sampling design.

https://doi.org/10.17221/141/2019-JFS

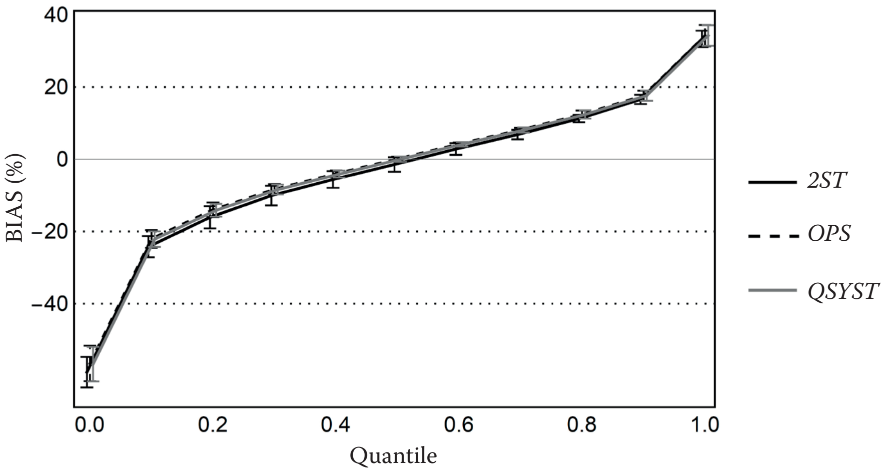

Figure 2. Quantile plot of relative bias in a predicted stand mean (bias is in percent of the actual mean). The error bars represent the interval of ± one standard deviation of the quantile mean across age classes and sample size levels. In case of *2ST*, the standard deviation includes variation over the design variants *2ST100*, *2ST60*, and *2ST30*

For the two-stage designs, the median bias in a predicted stand mean was –1.5% with a standard deviation of 2 percentage points across age classes, sample size levels, and design variants (*2ST100*, *2ST60*, and *2ST30*). Age class contributed most to the variation. Median bias for both *OPS* and *QSYST* was 0.0% with a standard deviation of 0.08% over age classes and sample size levels. Estimates of bias were seemingly independent of stand size (correlation < 0.02). The bias results are summarized in Figure 2. To note is an absence of effects related to sampling designs.

Among-stand variance

Under the two-stage designs, EBLUP estimates of the among-stand variance obtained from relatively small sample sizes (≤ 100) were, in general, greater than the actual variance and highly variable (Figure 3). This applies in particular to the first age class with the largest (positive) skewness in *mVOL* $m^3 \cdot ha^{-1}$. Under the two-stage designs, a minimum sample size of approximately 100 appears necessary to keep an absolute bias below 10%. Due to large standard errors, we failed to reject the null hypothesis (*t*-test) of equal bias in age class specific estimates from the three two-stage designs. However, over all age classes, and with the smallest (MIN) sample size level, the average bias of 23% in the estimates from *2ST30* was significantly higher than the bias of 11% and 12% in estimates from *2ST100* and *2ST60*. Overall, an increase in sample size of 1 unit lowered the bias by approximately 0.9 percentage points.

Under the *OPS* designs the retrieved estimates of the among-stand variance (with the modified analysis of variance) were, in age class two to seven, approximately 30% below the actual values. In age class one, however, the variance was overestimated by approximately 20%. The stronger skewness in this age class, and the stronger correlation between *mVOL* $m^3 \cdot ha^{-1}$ and *mCHT* creates a wider separation of the within age-class strata values of *mCHT* and, by extension, the residuals in *mVOL* $m^3 \cdot ha^{-1}$. These effects combine to generate an inflated estimate of variance. The small standard errors in retrieved estimates is a direct consequence of the averaging of the estimates obtained with 2, ..., 5 sorted residuals per cluster. Sample size level had no apparent effect on bias.

Retrieved among-stand variances obtained with the *QSYST* designs underestimated the actual variance by approximately 50% with no apparent effect of age class or sample size. As in the case of *OPS*, the standard error in a retrieved variance was much lower than with the *2ST* designs.

Coverage of confidence intervals for a stand mean

Coverage of nominal 95% confidence intervals for a stand mean was significantly improved by a provision of an estimate of the among-stand variance. Without such an estimate, the overall achieved coverage was 0.66 for the two-stage designs; with only minor (0.01) differences among the variants *2ST100*, *2ST60*, and *2ST30*, and sample size levels. With an estimate of the among-stand variance, the coverage increased to 0.82 (*2ST100*), 0.87 (*2ST60*), and 0.90 (*2ST30*). The graphics in Figure 4 emphasize the improvements in coverage with the provision of an estimate of the among-stand variance.

With the *OPS* design, coverage was dependent on sample size level (cf. Figure 3). Without a retrieved among-stand variance the coverage decreased from 0.82 to 0.68 as sample size level increased

https://doi.org/10.17221/141/2019-JFS

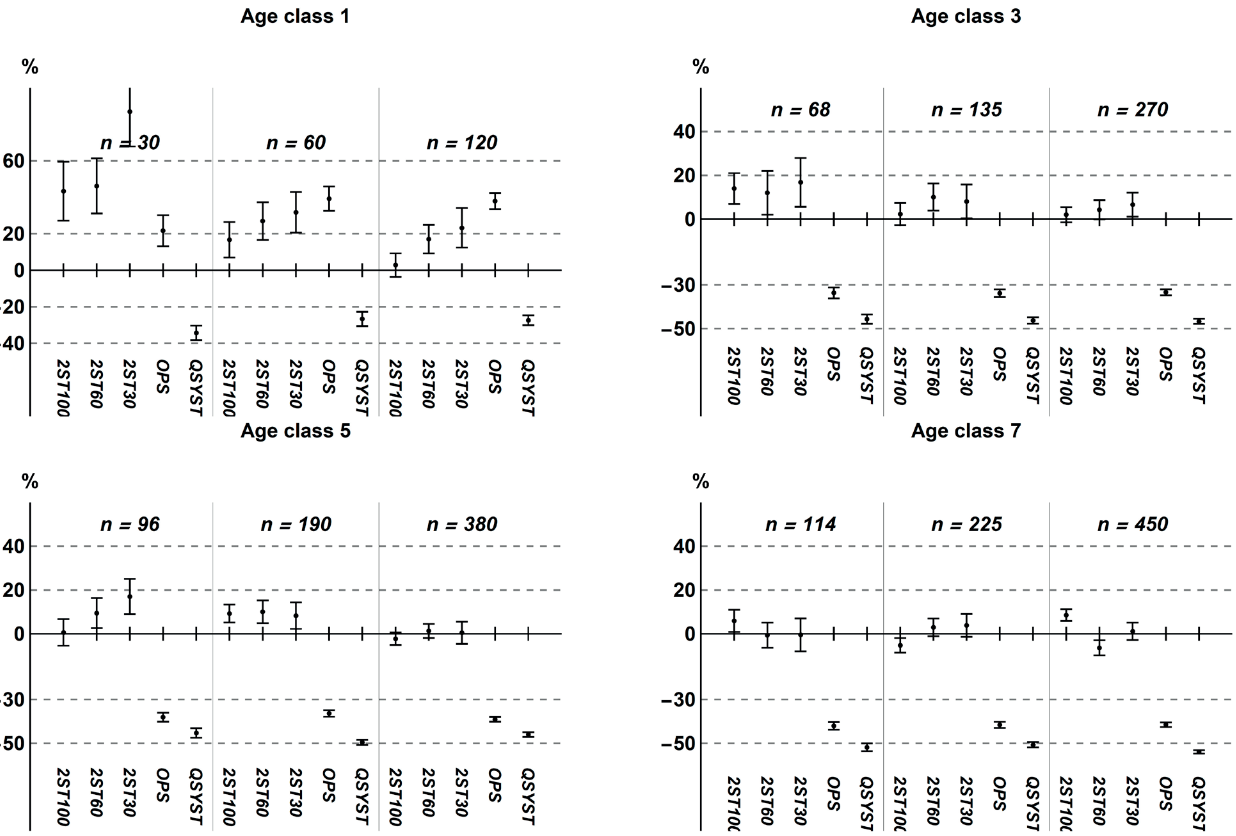

Figure 3. Relative difference between estimates and observations of the among-stand variance by design and sample size (The error bars represent the interval of ± one standard error of a relative difference in age class 1, 3, 5, and 7)

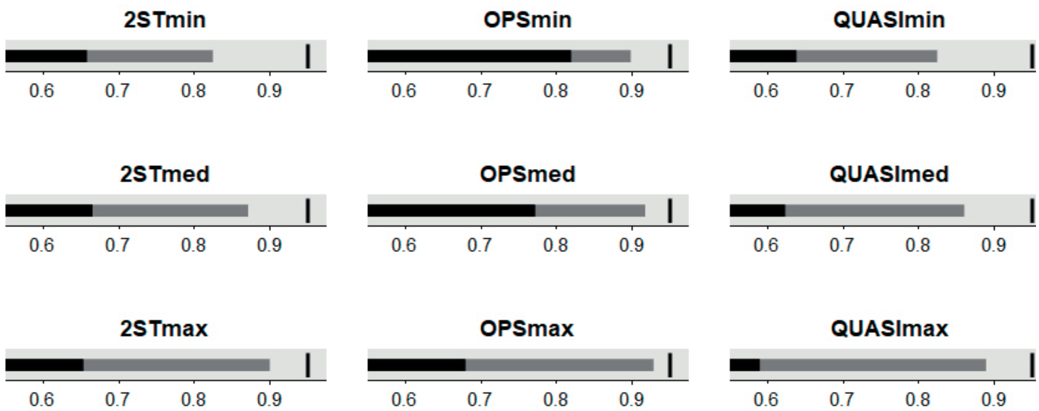

Figure 4. Coverage of 7 000 nominal 95% confidence intervals for a stand mean by design and sample size levels (black horizontal bar indicates the coverage without an estimate of the among-stand variance, gray bar indicates the improvement in coverage due to the provision of an estimate of the among-stand variance, the target coverage of 0.95 is indicated), *2ST* – two-stage sampling design, *OPS* – one-per-stratum, *QSYST* – quasi systematic sampling design

from MIN to MAX. The explanation comes from the faster decline in an estimate of errors in the regression coefficients with an increase in sample size compared to a smaller (if any) decline in the bias in a predicted stand mean. With a retrieved among-stand variance, the *OPS* coverage was improved, and it now increased with sample size level; from 0.90 (MIN), to 0.92 (MED), and to 0.93 (MAX).

https://doi.org/10.17221/141/2019-JFS

Coverage obtained under the *QSYST* design was more similar to results from the two-stage designs than to results from *OPS* (Figure 4). Without a retrieved among-stand variance, the coverage decreased from 0.64 to 0.59 as sample size increased from MIN to MAX. With a retrieved among-stand variance the *QSYST* coverage increased with sample size levels from 0.82 (MIN), to 0.86 (MED), and 0.89 (MAX).

Without an estimate of the among-stand variance, the coverage in age class one was, for all designs, approximately 0.20 lower than in older classes. Yet with estimates of the among-stand variance, the coverage became at par with coverage in older classes.

DISCUSSION

In forest enterprises with stand-oriented forest management practices, the provision of credible stand level predictions of forest resource attributes like stem wood volume is important. With the advancements in airborne laser scanning techniques and data processing (Corona, Fattorini 2008; Melville et al. 2015; Saarela et al. 2015), optical sensors (Holmström, Fransson 2003; Koch 2011), and unmanned airborne vehicles (Puliti et al. 2017), it is now possible to generate wall-to-wall model-based predictions of desired attribute values from a relatively small number of field sample plots selected according to a probability sampling design (Fattorini et al. 2009; Grafström, Ringvall 2013; Corona, 2016; Grafström et al. 2017).

From a modelling and efficiency perspective, it is advantageous to distribute the forest inventory samples uniformly across the population of interest (Mostafa, Ahmad 2017; Pagliarella et al. 2018; Räty et al. 2018) or to emulate the population distribution of one or more auxiliary variables (Grafström, Ringvall 2013; Grafström et al. 2017). These and other sampling approaches are becoming popular, and have in many instances replaced stand-level inventories (Duplat, Perrotte 1981; Mäkelä, Pekkarinen 2004; Muukkonen, Heiskanen 2007; Nothdurft et al. 2009; Kangas et al. 2018).

These advances have promoted accurate population-level estimates of forest resource attributes. However, these designs will be deficient for stand-level inference in forests with significant stand effects, if the design does not allow a direct estimation of the among-stand variance (Breidenbach et al. 2016; Magnussen, Breidenbach 2017). The deficiency is centered on poor estimates of the uncertainty in a predicted stand mean (Magnussen 2018). Deficient unless, of course, a reasonable estimate of the among-stand variance can be retrieved by other means (Breidenbach et al. 2016; Magnussen, Breidenbach 2017).

Estimators of an among-stand variance require at least two samples collected within a minimum number of stands. If there are stand effects, then it is well known that allocating more than one sample to a stand will lower the population level efficiency of a sampling design (Kleinn 1994; Tam 1995). We saw this in the lower efficiency of the two stage designs. Their efficiency declined – for a given overall sample size – with a decline in the number of stands in the sample. From a standpoint of population-level efficiency, sampling should be limited to at most two units per stand. We saw that it may be necessary to allocate two sample units to a relatively large number of stands (approx. 30) in order to get reasonably accurate estimates of the among-stand variance by traditional methods (e.g. EBLUP). For a given sample size, the bias in the among-stand variance estimate increased with a decrease in the number of stands with two sample units. Without a reasonable anticipated value of the among-stand variance component, it will not be possible to give more specific recommendations regarding the necessary number of stands to sample with two units (Maas, Hox 2005). Our results suggest that a two-stage design with two samples in each of approximately 30 stands can yield an age class (stratum) estimate of the among stand variance with an absolute bias of at most 10%.

Forest management may be better served by a combination of a sampling design that optimizes efficiency at the population or stratum level combined, and a method for retrieving the among-stand variance from data with – as a rule – one observation per stand represented in the sample. Our example with *OPS* and *QSYST* confirmed that these designs are not only more efficient for population (stratum) level inference, but they also generated analytical estimates of standard errors that were closer to the empirical (observed) estimates than the three variants of a two-stage design. Consequently, achieved population-level coverages were also better with *OPS* and *QSYST* than with the three *2ST* variants.

The proposed method, to retrieve an among-stand variance from sample data with at most one observation per stand, typically underestimated the

https://doi.org/10.17221/141/2019-JFS

among-stand variance by 30% to 50%. Nonetheless, the population and stand level coverage of nominal 95% confidence intervals with *OPS* was overall best, and the coverage with *QSYST* was at par with the coverage achieved with the three variants of a two-stage design. This may be serendipity specific for the studied population. Further research is needed to gauge how general this result is.

Our results also confirmed that merely a rough approximation of an among-stand variance is needed to significantly improve upon confidence intervals computed with the stand effects included in an estimate of the residual variance (Magnussen 2018). The proposed method is tailored to sampling designs that selects an ordered sample of one or more auxiliary variables where it is reasonable to assume that hidden random stand effects are sorted along a parallel gradient. The two parameters α1 and α2 required by our method will have to be estimated anew in each application. We used simulations with standard bivariate Gaussian variables for this purpose and anticipated levels of the reliability of a stand mean (the fraction of the total variance that is due to stands). In each case, estimates of the two parameters α1 and α2 are obtained by a standard minimization of the squared error in the pursued estimate of the (known) among-stand variance.

Our proposed new method – for retrieving an among-stand variance from sample data collected under a design that does not allow a direct estimation – was also superior to a methods proposed in an earlier study (Magnussen 2018). Under *OPS*, the absolute bias incurred with the new method was at most 50% of the bias with the older method (not shown). Under *QSYST*, the results for age classes one to three were similar, but in age classes four to seven, the absolute bias with the new and the older method was comparable to within 10%.

CONCLUSION

Field sampling designs with sample units uniformly distributed across space or feature space of important auxiliary variables are efficient, and they usually deliver accurate population and stratum estimates of forest resource attributes. However, they rarely allow a direct estimation of the among stand variance which is required for computing a reliable confidence interval for a predicted stand mean. The proposed new method for retrieval of this variance works for any design with sample locations stratified by an auxiliary variable. As demonstrated, confidence intervals for a stand mean obtained with the proposed method were better or at least at par with intervals obtained with less efficient designs that allow a direct estimation of the among-stand variance.

REFERENCES

Anderson T.W., Darling D.A. (1952): Asymptotic theory of certain „goodness of fit“ criteria based on stochastic processes. Annals of Mathematical Statistics, 23: 193–212.

Bailey R.L., Dell T.R. (1973): Quantifying diameter distributions with the Weibull function. Forest Science, 19: 97–104.

Bolker B.M., Brooks M.E., Clark C.J., Geange S.W., Poulsen J.R., Stevens M.H.H., White J.S.S. (2009): Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution, 24: 127–135.

Breidenbach J., Magnussen S., Rahlf J., Astrup R. (2018): Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. Remote Sensing of Environment, 212: 199–211.

Breidenbach J., McRoberts R.E., Astrup R. (2016): Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sensing of Environment, 173: 274–281.

Breidt F.J., Opsomer J.D., Sanchez-Borrego I. (2016): Nonparametric Variance Estimation Under Fine Stratification: An Alternative to Collapsed Strata. Journal of the American Statistical Association, 111: 822–833.

Cao Q.V. (2004): Predicting parameters of a Weibull function for modeling diameter distribution. Forest Science, 50: 682–685.

Casella G., Berger R.L. (2002): Statistical Inference. Pacific Grove, Duxbury Press: 660.

Cochran W.G. (1977): Sampling Techniques. New York, Wiley: 380.

Cordy C.B. (1993): An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. Statistics and Probability Letters, 18: 353–362.

Corona P. (2016): Consolidating new paradigms in large-scale monitoring and assessment of forest ecosystems. Environmental research, 144: 8–14.

Corona P., Fattorini L. (2008): Area-based lidar-assisted estimation of forest standing volume. Canadian Journal of Forest Research, 38: 2911–2916.

Corona P., Fattorini L., Franceschi S., Chirici G., Maselli F., Secondi L. (2014): Mapping by spatial predictors exploiting remotely sensed and ground data: A comparative design-based perspective. Remote Sensing of Environment, 152: 29–37.

https://doi.org/10.17221/141/2019-JFS

Dahlke M., Breidt F.J., Opsomer J.D., Van Keilegom I. (2013): Nonparametric endogenous post-stratification estimation. Statistica Sinica, 23: 189–211.

Dalenius T., Hodges J.L.J. (1959): Minimum variance stratification. Journal of the American Statistical Association, 54: 88–101.

Darroch N., Ratcliff D. (1971): A characterization of the Dirichlet distribution. Journal of the American Statistical Association, 66: 641–643.

Duane M.V., Cohen W.B., Campbell J.L., Hudiberg T., Turner D.P., Weyermann D.L. (2010): Implications of alternative field-sampling designs on Landsat-based mapping of stand age and carbon stocks in Oregon forests. Forest Science, 56: 405–416.

Dubey S.D. (1970): Compound gamma, beta and F distributions. Metrika, 16: 27–31.

Duplat P., Perrotte G. (1981): Inventory and Estimation of the Growth of Forest Stands. Paris, Office National des Forêts: 432.

Eerikäinen K. (2009): A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the Finnish National Forest Inventory. Forest Science, 55: 480–493.

Ene L.T., Gobakken T., Andersen H.E., Næsset E., Cook B.D., Morton D.C., Babcock C., Nelson R. (2018): Large-area hybrid estimation of aboveground biomass in interior Alaska using airborne laser scanning data. Remote Sensing of Environment, 204: 741–755.

Evert F. (1983): An equation for estimating total volume of both stands and single trees of black spruce. The Forestry Chronicle, 59: 26–29.

Fattorini L. (2015): Design-based methodological advances to support national forest inventories: a review of recent proposals. iForest - Biogeosciences and Forestry, 8: 6–11.

Fattorini L., Franceschi S., Pisani C. (2009): A two-phase sampling strategy for large-scale forest carbon budgets. Journal of Statistical Planning and Inference, 139: 1045–1055.

Fazar W. (1959): Program evaluation and review technique. The American Statistician, 13: 10–16.

Fischer M. (2010): Multivariate Copulae. In: Kurowicka D., Joe H. (eds): Dependence Modeling. Singapore, World Scientific: 19–36.

Fraley C., Raftery A.E. (1998): How many clusters? Which clustering method? Answers via model-based cluster analysis. Computer Journal, 41: 314–327.

Goerndt M.E., Monleon V.J., Temesgen H. (2011): A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. Canadian Journal of Forest Research, 41: 1189–1201.

Grafström A., Ringvall A.H. (2013): Improving forest field inventories by using remote sensing data in novel sampling designs. Canadian Journal of Forest Research, 43: 1015–1022.

Grafström A., Saarela S., Ene L.T. (2014): Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Canadian Journal of Forest Research, 44: 1156–1164.

Grafström A., Schnell S., Saarela S., Hubbell S.P., Condit R. (2017): The continuous population approach to forest inventories and use of information in the design. Environmetrics, 28: e2480-n/a.

Harvey A.C. (1976): Estimating regression models with multiplicative heteroscedasticity. Econometrica, 44: 461–465.

Holmström H., Fransson J.E.S. (2003): Combining remotely sensed optical and radar data in kNN-estimation of forest variables. Forest Science, 49: 409–418.

Izenman A.J. (1991): Recent development in nonparametric density estimation. Journal of the American Statistical Association, 86: 205–224.

Jiang J., Lahiri P. (2006): Mixed Model Prediction and Small Area Estimation. TEST, 15: 1–96.

Junttila V., Karuanne T., Leppänen V. (2010): Estimation of forest stand parameters from airborne laser scanning using calibrated plot databases. Forest Science, 56: 257–270.

Kangas A., Astrup R., Breidenbach J., Fridman J., Gobakken T., Korhonen K.T., Maltamo M., Nilsson M., Nord-Larsen T., Næsset E. (2018): Remote sensing and forest inventories in Nordic countries – roadmap for the future. Scandinavian Journal of Forest Research, 33: 397–412.

Katila M., Tomppo E. (2002): Stratification by ancillary data in multisource forest inventories employing k-nearest-neighbour estimation. Canadian Journal of Forest Research, 32: 1548–1561.

Kleinn C. (1994): Comparison of the performance of line sampling to other forms of cluster sampling. Forest Ecology and Management, 68: 365–373.

Koch B. (2011): Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote Sensing, 65: 581–590.

Levene H. (1960): Robust tests for equality of variances. In: Olkin I. (Ed.): Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Standford, Standford University Press: 278–292.

Li F., Zhang L., Davis C.J. (2002): Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. Forest Science, 48: 47–58.

Liu G., Liang K.Y. (1997): Sample Size Calculations for Studies with Correlated Observations. Biometrics, 53: 937–947.

Maas C.J.M., Hox J.J. (2005): Sufficient Sample Sizes for Multilevel Modeling. Methodology, 1: 86–92.

https://doi.org/10.17221/141/2019-JFS

MacLeod D. (1978): The Forest Management Institute Tree Data Bank. Ottawa, Forest Management Institute in Ottawa. Information Report FMR-X-112: 16.

Magnussen S. (1986): Diameter distributions in Picea abies described by the Weibull model. Scandinavian Journal of Forest Research, 1: 493–502.

Magnussen S. (2016): A new mean squared error estimator for a synthetic domain mean. Forest Science, 63: 1–9.

Magnussen S. (2018): An estimation strategy to protect against over-estimating precision in a LiDAR-based prediction of a stand mean. Journal of Forest Science, 64: 497–505.

Magnussen S., Breidenbach J. (2017): Model-dependent forest stand-level inference with and without estimates of stand-effects. Forestry: An International Journal of Forest Research, 90: 675–685.

Magnussen S., Frazer G., Penner M. (2016): Alternative mean-squared error estimators for synthetic estimators of domain means. Journal of Applied Statistics, 43: 2550–2573.

Magnussen S., Mauro F., Breidenbach J., Lanz A., Kändler G. (2017): Area-level analysis of forest inventory variables. European Journal of Forest Research, 136: 839–855.

Magnussen S., Næsset E., Gobakken T. (2015): LiDAR supported estimation of change in forest biomass with time invariant regression models. Canadian Journal of Forest Research, 45: 1514–1523.

Mäkelä H., Pekkarinen A. (2004): Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecology and Management, 196: 245–255.

Mauro F., Molina I., García-Abril A., Valbuena R., Ayuga-Téllez E. (2016): Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics, 27: 225–238.

Mbachu H., Nduka E., Nja M. (2012): Designing a Pseudo R-Squared Goodness-of-Fit Measure in Generalized Linear Models. Journal of Mathematics Research, 4: 148.

Melville G., Stone C., Turner R. (2015): Application of LiDAR data to maximise the efficiency of inventory plots in softwood plantations. New Zealand Journal of Forestry Science, 45: 1–9.

Mostafa S.A., Ahmad I.A. (2017): Recent developments in systematic sampling: A review. Journal of Statistical Theory and Practice, 12: 1–21.

Muukkonen P., Heiskanen J. (2007): Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories. Remote Sensing of Environment, 107: 617–624.

Næsset E. (2014): Area-based inventory in Norway – from innovation to an operational reality. In: Maltamo M., Naesset E., Vauhkonen J. (eds) :Forestry Applications of Airborne Laser Scanning. Dordrecht, Springer: 215–240.

Newton P., Amponsah I. (2007): Comparative evaluation of five height–diameter models developed for black spruce and jack pine stand-types in terms of goodness-of-fit, lack-of-fit and predictive ability. Forest Ecology and Management, 247: 149–166.

Nothdurft A., Saborowski J., Breidenbach J. (2009): Spatial prediction of forest stand variables. European Journal of Forest Research, 128: 241–251.

Pagliarella M.C., Corona P., Fattorini L. (2018): Spatially-balanced sampling versus unbalanced stratified sampling for assessing forest change: evidences in favour of spatial balance. Environmental and Ecological Statistics, 25: 111–123.

Payandeh B. (1991): Plonski's (metric) yield tables formulated. The Forestry Chronicle, 67: 545–546.

Pinheiro J.C., Bates D.M. (2000): Mixed-effects models in S and S-plus. New York, Springer: 1–528.

Plonski W. (1960): Normal Yield Tables for Black Spruce, Jack Pine, Aspen, White Birch, Tolerant Hardwoods, White Pine, and Red Pine for Ontario. Ontario, Ontario Department od Land and Forests. Silvicultural Series Bulletin 2: 39.

Puliti S., Ene L.T., Gobakken T., Næsset E. (2017): Use of partial-coverage UAV data in sampling for large scale forest inventories. Remote Sensing of Environment, 194: 115–126.

Räty M., Heikkinen J., Kangas A. (2018): Assessment of sampling strategies utilizing auxiliary information in large-scale forest inventory. Canadian Journal of Forest Research, 48: 749–757.

Rennolls K., Wang M. (2005): A new parameterization of Johnson's SB distribution with application to fitting forest tree diameter data. Canadian Journal of Forest Research, 35: 575–579.

Saarela S., Grafström A., Ståhl G., Kangas A., Holopainen M., Tuominen S., Nordkvist K., Hyyppä J. (2015): Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. Remote Sensing of Environment, 158: 431–440.

Saarela S., Holm S., Grafström A., Schnell S., Næsset E., Gregoire T.G., Nelson R.F., Ståhl G. (2016): Hierarchical model-based inference for forest inventory utilizing three sources of information. Annals of Forest Science, 73: 895–910.

Särndal C.E., Swensson B., Wretman J. (1992): Model Assisted Survey Sampling. New York, Springer: 694.

Self S.G., Mauritsen R.H. (1988): Power/Sample Size Calculations for Generalized Linear Models. Biometrics, 44: 79–86.

Sexton J.O., Bax T., Siqueira P., Swenson J.J., Hensley S. (2009): A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeast-

https://doi.org/10.17221/141/2019-JFS

ern North America. Forest Ecology and Management, 257: 1136–1147.

Snedecor G.W., Cochran W.G. (1971): Statistical Methods. Iowa, Iowa State University Press: 593.

Snijders T.A. (2005): Power and Sample Size in Multilevel Linear Models. In: Everitt B.S., Howell D.C. (eds): Encyclopedia of Statistics in Behavioral Science, 3: 1570–1573.

Spurr S.H. (1952): Forest Inventory. New York, Ronald Press: 476.

Stevens D.L., Olsen A.R. (2004): Spatially balanced sampling of natural resources. Journal of the American Statistical Association, 99: 262–278.

Tam S.M. (1995): Optimal and robust strategies for cluster sampling. Journal of the American Statistical Association, 90: 379–382.

Tomppo E., Malimbwi R., Katila M., Mäkisara K., Henttonen H.M., Chamuya N., Zahabu E., Otieno J. (2014): A sampling design for a large area forest inventory: Case Tanzania. Canadian Journal of Forest Research, 44: 931–948.

von Lüpke N., Hansen J., Saborowski J. (2012): A three-phase sampling procedure for continuous forest inventory with partial re-measurement and updating of terrestrial sample plots. European Journal of Forest Research, 131: 1979–1990.

Wilhelm M., Tillé Y., Qualité L. (2017): Quasi-systematic sampling from a continuous population. Computational Statistics & Data Analysis, 105: 11–23.

Received: December 4, 2019
Accepted: April 3, 2020