Can. J. Remote Sensing, Vol. 38, No. 5, pp. 644–654, 2012

Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables

Jacob Strunk, Hailemariam Temesgen, Hans-Erik Andersen, James P. Flewelling, and Lisa Madsen

Abstract. Using lidar in an area-based model-assisted approach to forest inventory has the potential to increase estimation precision for some forest inventory variables. This study documents the bias and precision of a model-assisted (regression estimation) approach to forest inventory with lidar-derived auxiliary variables relative to lidar pulse density and the number of sample plots. For managed forests on the Lewis portion of the Lewis-McChord Joint Base (35 025 ha, 23 290 forested) in western Washington state, we evaluated a regression estimator for combinations of pulse density (.05–3 pulses/m^2) and sample size (15–105 plots) to estimate five forest yield variables: basal area, volume, biomass, number of stems, and Lorey's height. The results indicate that there is almost no loss in precision in using as few as .05 pulses/m^2 relative to 3 pulses/m^2. We demonstrate that estimation precision declined quickly for reduced sample sizes (as expected from sampling theory); but of more importance we demonstrate that sample size has a dramatic effect on the validity of inferences. Our investigations indicate that for our test dataset that central limit theorem based confidence intervals were too small on average for sample sizes smaller than 55. The results from this study can aid in identifying design components for forest inventory with lidar which satisfy users' objectives.

Résumé. L'utilisation du lidar pour la réalisation des inventaires forestiers en utilisant une approche basée sur la modélisation de la surface permet d'améliorer la précision des estimations pour certaines variables des inventaires forestiers. Dans cette étude, on documente le biais et la précision d'une approche basée sur la modélisation (estimation par régression) pour les inventaires forestiers utilisant des variables auxiliaires dérivées par lidar relativement à la densité de l'impulsion lidar et au nombre de placettes-échantillons. Dans le cas des forêts aménagées situées dans la portion Lewis du territoire de la Lewis-McChord Joint Base (35 025 ha, 23 290 en forêt), dans l'ouest de l'état de Washington, on a évalué un estimateur par régression pour des combinaisons de densités d'impulsions (,05–3 impulsions/m^2) et de tailles d'échantillons de (15–105 parcelles) pour estimer cinq variables du rendement forestier, i.e. la surface terrière, le volume, la biomasse, le nombre de tiges et la hauteur de Lorey. Les résultats montrent qu'il y a très peu de perte de précision en utilisant aussi peu que ,05 impulsion/m^2 par rapport à 3 impulsions/m^2. On démontre que la précision de l'estimation diminue rapidement dans le cas des échantillons à dimension réduite (comme on peut s'y attendre en fonction de la théorie d'échantillonnage) – mais plus important encore, on démontre que la taille de l'échantillon a un effet dramatique sur la validité des inférences. Nos recherches indiquent que, pour notre ensemble de données de référence, les niveaux de confiance basés sur le théorème central limite étaient trop faibles en moyenne pour les échantillons de taille inférieure à 55. Les résultats de cette étude peuvent aider à l'identification des composantes dans la conception des inventaires forestiers basés sur le lidar permettent ainsi de rencontrer les objectifs des utilisateurs.
[Traduit par la Rédaction]

Introduction

Over the last several decades many studies have used models to relate area-based measurements of forest attributes to summary statistics calculated from airborne scanning lidar for the same areas (lidar metrics) including basal area (Lefsky et al., 1999), stem volume (Næsset, 1997b), stand height (Næsset, 1997a), biomass (Lefsky et al., 1999; Means et al., 1999), and others. The many studies documenting the area-based relationship between lidar and forest variables are important because they indicate that there is a strong association between the two types of variables, although the studies may provide limited descriptions of how the models should be used for practical

Received 19 April 2012. Accepted 30 August 2012. Published on the Web at http://pubs.casi.ca/journal/cjrs on 20 December 2012.

Jacob Strunk[1] and Hailemariam Temesgen. Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, Oregon, USA.

Hans-Erik Andersen. USDA Forest Service, PNW Research Station, Seattle, WA, USA.

James P. Flewelling. Seattle Biometrics and Analysis, Seattle, WA, USA.

Lisa Madsen. Department of Statistics, Oregon State University, Corvallis, Oregon, USA.

[1]Corresponding author (e-mail: Jacob.Strunk@oregonstate.edu).

inference. The most common approach to inference in forest inventory and survey sampling in general is design-based inference. In the design-based context a probability sample is used to make inference about parameters of a finite population (e.g., the population mean or total). Lidar metrics and a model can be leveraged in a design approach by using a model-assisted approach (Parker and Evans, 2007; Corona and Fattorini, 2008).

As with planning a traditional design-based inventory, it is desirable to quantify the effect that inventory design parameters have on estimation precision. Increasing the number of ground plots, for example, is one possible way to modify a design to increase estimation precision. In the case of model-assisted forest inventory with lidar, lidar pulse density is an important consideration in addition to the number of ground plots. Lidar pulse density has been examined in a forestry setting using at least three approaches, including thinning the data (Holmgren, 2004; Gobakken and Næsset, 2008; Maltamo et al., 2006; Magnussen et al., 2010), performing multiple lidar acquisitions for the same area (Parker and Glass, 2004; Næsset, 2004), and generating synthetic lidar data (Lovell et al., 2005). There appears to be a consensus among findings that reduction in pulse density increases the variability in lidar metrics (sometimes this is expressed implicitly in studies, such as in increased standard error (SE) estimates for model coefficients for reduced pulse density). This is reasonable given that most of the lidar-derived predictors are consistent statistics, thus we would expect their sampling variation to decline as the number of lidar points increases. However, in part due to methodological differences, studies differ in the rate at which reduced pulse density affects the residual variability of a fitted model. The study by Gobakken and Næsset (2008) additionally details the effects of number of ground plots and the size of ground plots; both of these had an effect on residual bias and root mean square error (RMSE).

The study by Magnussen et al. (2010) on the outset appears similar to our study, but their objectives, methods, inferences, and the theoretical basis for their inferences (model-based) are all different from our own. Because they used a model-based perspective to explore replication effects and made inferences to repeated measurements in time, we have chosen not to delve into any comparisons with their inferences here. Also, as all of their results concern a nonlinear combination of predictors we cannot easily discuss their results concerning the sampling distributions of lidar metrics or estimators to other studies.

A limitation in attempting to use the results from previous studies to guide practical applications is that they focus on the effects of pulse density and sample size on the behavior of lidar metrics, on the behavior of fitted model coefficients, and on the variability of residuals. They do not generally directly assess the practical implications of sample size and pulse density on inference regarding forest inventory applications when lidar-based models are used to assist in estimation. Some of the implications are fairly obvious with respect to estimation given the theoretical properties of the estimators. This is especially the case for pulse density, given that most studies agree that using even very low pulse densities has little effect on modeling. However, we assert that it is crucial in the literature to directly link implications about estimation to pulse density and sample size to facilitate understanding and adoption. It is also important to consider the viability of estimation strategies for different pulse densities and sample sizes given that the estimators used for estimation with lidar are only asymptotically valid. For example, an important consideration that will not be obvious from previous studies is that the ability to develop valid confidence intervals is highly dependent upon sample size. In this case the guideline of having at least 30 observations is insufficient to obtain reasonably valid confidence intervals. Investigations of viability are clearly important, because a number of previous lidar studies used numbers of observations below the threshold for which we found viable design-based inference to be feasible.

The first objective of this study was to quantify the effects of lidar pulse density and sample size on design-based estimation precision with auxiliary lidar. We then evaluated the influence of pulse density and sample size on our ability to make valid inferences from our estimates. Specifically, did our mean estimator unbiasedly estimate the population mean, and did confidence bounds cover the true mean with the specified level of confidence. These aspects of inference were addressed through the use of resampling methods. We simulated sampling distributions of model-assisted mean and mean variance estimators for different combinations of pulse density and sample size and compared the sampling distributions with the corresponding estimators. The simulations were performed treating our sample as the population which prevents us from making inference about the entire forest. However, this did not affect our ability to make inference about the sampling properties of the estimators. By treating our sample as the population we could still make the same inferences about the estimation strategy that could be made if we were estimating the mean of the entire original population, but by treating our sample as the population we could simplify our presentation. We describe how our estimator behaves for forest attributes such as volume including whether the estimator is unbiased and whether 95% confidence intervals cover the true mean 95% of the time as they should. This will guide us in developing forest inventories with lidar that can support correct management inferences. The related study by Strunk et al. (2012) provides a practical demonstration in which the same data were used to make inference about the entire population using the same set of estimators.

Methods

Study site

This study was conducted on the forested areas of the Fort Lewis military installation (47° 30′ 58″ N 122° 35′ 11″ W)

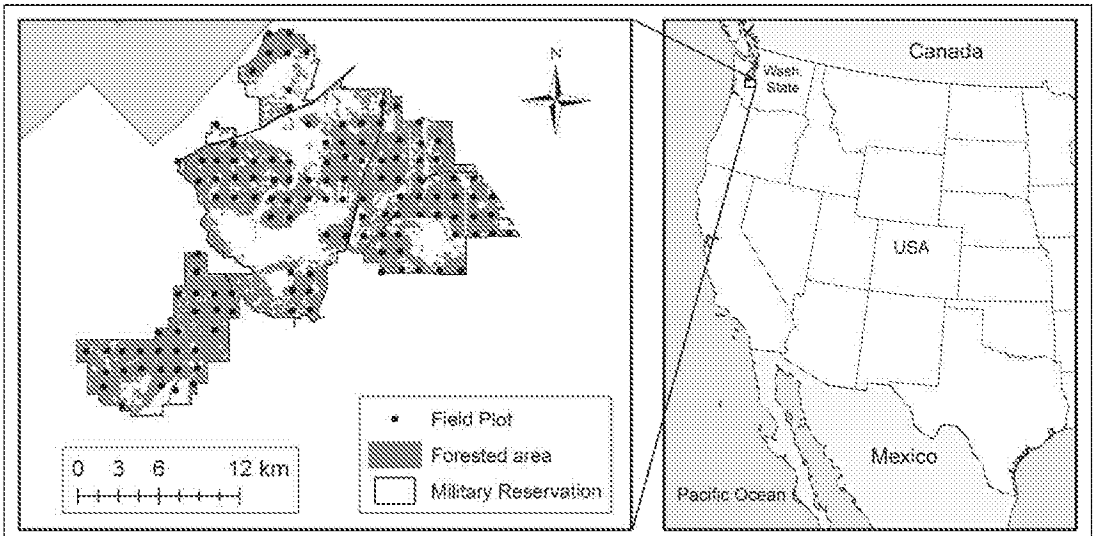

Figure 1. Location of the study area in central western Washington state.

in western Washington state, USA (**Figure 1**). Managed forests are dominated by conifer trees, especially Douglas-fir with interspersed western hemlock (*Tsuga heterophylla (Raf.)* Sarg.), western red cedar (*Thuja plicata* Donn ex D. Don), and very limited amounts of Sitka spruce (*Picea sitchensis* (Bong.) Carrièr) and pacific yew (*Taxus brevifolia* Nutt.). Hardwood trees are also present on managed sites, mostly in the wet areas, including red alder (*Alnus Rubra* Bong.), big leaf maple (*Acer Macrophyllum* Pursh), and black cottonwood (*Populus balsamifera ssp. Trichocarpa* (Torr. & A. Gray ex Hook.) Brayshaw).

Data description

Forest inventory variables

Values of forest inventory variables were calculated for 120 circular 809 m^2 (0.2 acres) permanent (plot centers and reference trees are clearly monumented and detailed navigation guides are available for each plot) forest inventory plots measured between May of 2004 and January of 2005. The plots are distributed on a 1.28 × 1.28 km grid within forested areas of the study site. Field measurements recorded for each tree that we used included diameter at breast height (dbh) and species. Because measurements were not consistently taken on trees less than 20.32 cm in diameter, trees with a dbh less than 20.32 cm were excluded from plot tabulations. Tree heights were measured for a subset of trees on plots, approximately 2.5 trees per plot.

Unmeasured Douglas-fir heights were estimated using a nonlinear mixed-effects model (Temesgen et al., 2008) developed specifically for this study using Fort Lewis inventory data. There were insufficient numbers of trees from other species to model heights in this way. Heights for other species were estimated using the Wykoff et al. (1982) model form with Eugene area (Oregon state, USA) parameter estimates from the Pacific Northwest Coast Variant Overview for the Forest Vegetation Simulator (Keyser, 2010). The values for five variables including basal area (ba), total stem volume including top and stump (vol), Lorey's height (lor, mean height weighted proportional to ba), total aboveground biomass (bm), and number of trees (stems) were calculated for each plot (**Table 1**). Vol was estimated using the allometric models included with the National Volume Estimator Library plugin for Excel (USFS, 2008) and bm was estimated using the models provided by Jenkins et al. (2004).

Precision plot positioning

Coordinates for plot centers were collected between September 2007 and May 2008. Survey-grade differential global navigation satellite system (GNSS) receivers were used to survey plot centers marked with stakes. GNSS data were post-processed and differentially corrected using the Ensemble processing suite from Javad Navigation Systems (now Javad GNSS) to achieve the desired level of accuracy, approximately 1 m horizontal RMSE, under dense northwest United States forest canopy (Andersen et al., 2009; Clarkin, 2007).

Lidar data and variables

The lidar data used for this study were acquired in September 2005 (leaf on) within one growing season from

Table 1. Summary of plot level variables.

Variable	Mean	Mininum	Maximum	Standard deviation
ba (m^2/ha)	34.2	0.0	84.3	19.3
vol (m^3/ha)	437.0	0.0	1462.0	267.9
stems (ha)	220.5	0.0	864.9	156.9
bm (kg/ha)	281 214.3	0.0	914 784.5	172 487.1
lor (m)	34.9	0.0	56.7	9.2

the time of measurements. Lidar data were collected from a fixed wing aircraft. The airplane flying height was approximately 1000 m above ground, the scan angle was +/− 14 degrees, and the laser beam divergence was 0.3 mrad. The pulse repetition frequency (PRF) was approximately 71 kHz and the nominal pulse density was 4 pulses/m^2. The vendor collected the data with an Optech ALTM 3100.

The lidar metrics used in this study are a series of statistics calculated on the lidar height data. The statistics were computed on first return lidar heights above 1 m, including the mean, standard deviation, and percentiles (e.g., ht_5 and ht_{95} represent 5th and 95th percentile heights, respectively). We also computed the ratio of the number of first returns above a given height relative to the total number of first returns for a plot (e.g., $cover_1$ and $cover_2$ are the proportions of returns above 1 m and 2 m, respectively). Lidar heights were estimated by subtracting ground elevations from lidar elevations. Digital terrain modeling was performed using the algorithm described by Kraus and Pfeifer (1998) and implemented in FUSION (McGaughey, 2012) which appears to be comparable with the more common Axelsson (2000) approach in wooded areas (Sithole and Vosselman, 2004). Lidar metrics were also computed using FUSION software.

Regression estimation

In this study we used a regression estimator (Särndal et al., 1992) for the mean

$$\hat{\mu}_y = \mu_X^T \hat{\beta} \quad (1)$$

where $\hat{\mu}_y$ is the regression estimate of the response mean, $\hat{\beta}$ is a vector of regression coefficients fitted with ordinary least squares (OLS) including the intercept, and μ_X^T is the transpose of a vector of means of auxiliary variables including a 1.0 corresponding to the intercept.

This formulation of a regression estimator is appropriate when the auxiliary variables are available for the entire finite population of interest and when the sample was collected using an SRS design (Lohr, 1999; Gregoire and Valentine, 2008; Särndal et al., 1992) or another design where no weights are necessary. Other designs may require a slightly different formulation of Equation (1), which includes a correction term (an estimate of the bias) that is a function of the sampling weights and regression residuals. In many cases, as here, the estimate of bias is zero.

An accompanying standard error (SE) for Equation (1) is (Särndal et al., 1992)

$$\mathrm{SE}_{\hat{\mu}} \approx \sqrt{\frac{s_{\mathrm{reg}}^2}{n}} \quad (2)$$

where $\mathrm{SE}_{\hat{\mu}}$ is the SE of the regression estimate of the mean, n is the size of the sample, and s_{reg}^2 is the variance of the regression residuals with the denominator corrected for the number of parameters estimated in the model.

We excluded the finite population correction term that is present in the original because our simulation analysis was performed with replacement. Equation (2) is a simplified form of the estimator that does not account for variability in the explanatory variables. However we found, as Särndal et al. (1992) demonstrated, that there was no practical difference in estimates when the simplified estimator is used.

Using estimators appropriate for a simple random sample (SRS) for a systematic sample will yield unbiased mean estimates, but SEs under a systematic design will tend to be conservative (too large) when using SRS estimators unless periodicity of the attribute of interest matches the interval of the systematic sample. This bias of the variance estimator is typically ignored because there is not a design-unbiased estimator for systematic designs.

Simulations

General description

In this study we used resampling simulations to approximate the sampling distributions of estimators for combinations of pulse densities and sample sizes. The use of simulations is well established in the statistical literature, and theoretical bases for their use can be found in a variety of textbooks (Efron and Tibshirani, 1993; Wolter, 2007). Five thousand simulations were used to approximate the sampling distributions for each of the examined combinations. For convenience we defined our population as our complete sample, all 120 observations collected in the field. This approach is similar to the approach used by Andersen and Breidenbach (2007) to contrast several approaches to estimation with lidar. We elected to treat our sample as the population because we were not attempting to estimate parameters of the original population; we were attempting to describe general properties of an estimation strategy relative to pulse density and sample size. Our results were not the same as what would be obtained from another sample or population, but they can serve to demonstrate general trends and issues and can be indicative (rather than conclusive) of behavior for similar areas and sampling designs.

Lidar thinning

To assess the effect of pulse density on inference in model-assisted estimation the original lidar point cloud was thinned repeatedly for reduced pulse densities. The pulse densities examined were .05, .3, .45, .6, .85, 1, 2, and 3 pulses/m^2. Thinning, along with other lidar processing, was performed using FUSION. The thinning algorithm is a Bernoulli sampling algorithm, where the probability of including a particular point is determined by the desired point density on a plot relative to the observed point density. The first step in the thinning procedure was to thin the original lidar dataset for large (relative to the plots) 314 m wide squares centered on the ground plots. The large squares contain both the plots and buffer regions around the plots to eliminate edge effects within field plots when creating digital elevation models (DEMs) and thinning the lidar. Reduced pulse density data on the 314 × 314 m squares were then used to select ground returns and create bare-earth digital elevation models. A 3 m median smoother was applied to each ground DEM. DEM elevations were then subtracted from lidar elevations to calculate lidar height values. The reduced density elevation-corrected lidar data were then clipped to the extents of the precisely geo-referenced ground plots. Finally, lidar metrics were calculated for each thinned lidar dataset.

OLS regression

The relationship between measures of forest yield (e.g., ba) and lidar metrics was modeled using OLS regression. OLS models were fitted using a two-step procedure: (*i*) prior to the simulations predictor variables were selected to include in the models, and (*ii*) model parameters were re-estimated using the fixed set of predictors for each of 5000 random resamples. Model RMSE (or $\sqrt{s^2_{\mathrm{reg}}}$) values were calculated for each of the 5000 models (fit to the 5000 resamples of data) using both the entire population (σ_ε) and only the observations in resample used to fit the model (RMSE). We initially considered inclusion of an automated variable selection algorithm in the simulations to allow different variables for each simulation, but found that the choice of model-selection algorithm highly influenced the results.

The sets of predictor variables used for regression were selected by comparing the RMSEs of a variety of models using all 120 observations and comparing their Bayesian information criterion (BIC) scores, one of several common bases for variable selection (Schwarz, 1978). Additionally, we preferred simple models that included at least one height quantile metric and one cover metric. Interactions between variables selected for inclusion were also explored. No transformations of the responses were used because there was only very mild heteroskedasticity, and model-assisted inference does not require equal variance. The same set of predictors was used for ba, vol, stems, and bm (ht_{25}, $\mathrm{cover}_{1.37}$, and $\mathrm{cover}_{1.37} \times \mathrm{ht}_{70}$) with a slightly different set used for lor (ht_{70}, ht_{95}, $\mathrm{cover}_{1.37}$, and $\mathrm{cover}_{1.37} \times \mathrm{ht}_{95}$).

Unfortunately, model selection can result in overly optimistic inferences "especially so when the pool of candidate variables is large" (Shen et al., 2004). In our case we attempted to protect ourselves from this optimism by combining automated model selection with expert knowledge. This reduces the chance that a best model is selected due to an artifact in the data which may only effectively represent a trend present in the sample. Fortunately, for our sample it appears that many different related sets of predictors (similar types of variables) fared comparably in terms of BIC, RMSE, and simplicity; therefore, the selection of one from the many comparable sets of variables is less likely to represent an artifact of our sample.

Estimator performance

As previously mentioned, we used resampling simulations to evaluate estimators for different combinations of pulse density and sample size. We developed simulation distributions for estimators for each combination of pulse density and sample size using 5000 samples. For each simulation we first sampled from our population of 120 elements to a reduced sample size with replacement. For each sample we estimated the population mean using the regression estimator in Equation (1) and calculated deviations (or residuals) of mean estimates from the known population mean

$$\varepsilon_{\hat{\mu},i} = \hat{\mu}_i - \mu \tag{3}$$

The empirical distribution of mean residuals was evaluated using the mean (or bias)

$$B_{\hat{\mu}} = \frac{\sum_{i=1}^{5000} \varepsilon_{\hat{\mu},i}}{5000} \tag{4}$$

and standard deviation

$$\sigma_{\hat{\mu}} = \sqrt{\frac{\sum_{i=1}^{5000} \left(\varepsilon_{\hat{\mu},i} - B_{\hat{\mu}}\right)^2}{5000 - 1}} \tag{5}$$

A smaller bias and standard deviation of mean residuals indicates superior performance for a given configuration of pulse density and sample size.

In addition to the mean, we also calculated the $\mathrm{SE}_{\hat{\mu}}$ for each resample. The $\mathrm{SE}_{\hat{\mu}}$ values were used to evaluate $\mathrm{SE}_{\hat{\mu}}$ as an estimator of $\sigma_{\hat{\mu}}$, which we treat as a population parameter.

$\mathrm{SE}_{\hat{\mu}}$ values for each sample were differenced from $\sigma_{\hat{\mu}}$

$$\varepsilon_{\hat{\sigma}} = \mathrm{SE}_{\hat{\mu}} - \sigma_{\hat{\mu}} \tag{6}$$

Again, a small bias

$$B_{\hat{\sigma}} = \frac{\sum_{i=1}^{5000} \varepsilon_{\hat{\sigma},i}}{5000} \tag{7}$$

and small variance

$$\sigma_{\hat{\sigma}} = \sqrt{\frac{\sum_{i=1}^{5000} \left(\varepsilon_{\hat{\sigma},i} - B_{\hat{\sigma}}\right)^2}{5000 - 1}} \tag{8}$$

of residuals indicates superior performance for the analytical variance estimator. Because the number of samples is large, 5000, the variability in $\sigma_{\hat{\mu}}$ will be very small and this source of variation was ignored in Equations (6–8).

The sampling distributions of $\mathrm{SE}_{\hat{\mu}}$ values were also used to look at coverage probability, which is the proportion of confidence intervals that cover the population mean. Confidence limits were based upon the student's *t* distribution with $n-p$ degrees of freedom, where p is the number of parameters fitted in the regression model.

For each combination of pulse density and sample size we also calculated the relative precision (RP)

$$\mathrm{RP}_{n_k} = \frac{\sigma^2_{\hat{\mu}}}{\sigma^2_{\mathrm{SRS},120}} \tag{9}$$

where RP_{n_k} is the RP of the regression estimator for n_k observations and $\sigma^2_{\mathrm{SRS},120}$ is the variance of 5000 SRS sample mean estimates for 120 plots.

Relative precision is analogous to design effect (Särndal et al., 1992), which is used to roughly indicate the efficiency (precision relative to sample size) of an estimation strategy. Multiplication of relative precision reported in this study by the ratio $n_k/120$ will yield a value that is equivalent to design effect for a given pulse density and sample size.

Results and discussion

Modeling effects

The magnitudes of variability in residuals (σ_ε) differed by response variable, but the general behavior relative to pulse density and sample size was similar for the five response variables examined. Reduction in sample size caused a noticeable increase in residual variability (**Figure 2a**). Median σ_ε and median RMSE from 5000 simulations are shown for all of the variables for a fixed pulse density in **Table 2**. The shaded curves in **Figure 2** are empirical densities created from 5000 simulations for each of the combinations of pulse density and sample size. The effect of sample size was slight for samples with 45 or more observations. With fewer than 45 observations the median of the simulation distribution of model RMSEs and the widths of the simulation distributions both increased from less than 1% in moving from 55 to 45 observations to 7% in moving from 25 to 15 observations. This is the result of

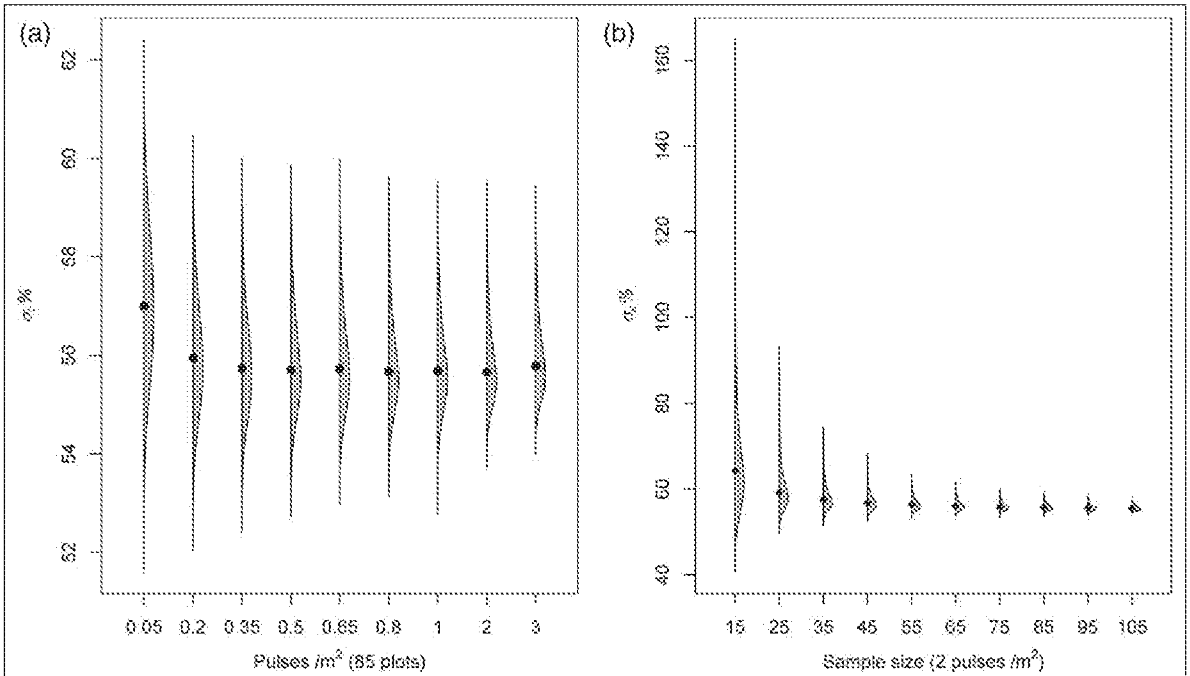

Figure 2. Simulation distributions of model RMSEs in percent of the mean by (a) pulse density and (b) sample size for stems. Black dots indicate the median RMSEs and the horizontal lines indicate bounds of 95% empirical confidence intervals.

Table 2. Residual summary statistics, σ_ε%, and RMSE %, for models fitted for reduced sample sizes using lidar with 1 pulse/m^2.

	Sample size									
Variable	15	25	35	45	55	65	75	85	95	105
ba (m^2/ha)	30.3 (21.3)	28.1 (23.4)	27.3 (24.3)	26.9 (24.8)	26.7 (25)	26.6 (25.3)	26.5 (25.4)	26.4 (25.5)	26.3 (25.6)	26.3 (25.7)
stems (ha)	64.2 (42.8)	59.1 (48.7)	57.5 (50.8)	56.8 (52.3)	56.3 (52.9)	56 (53.3)	55.8 (54)	55.7 (54.1)	55.6 (54.3)	55.5 (54.6)
vol (m^3/ha)	36.5 (24.1)	34.1 (27.4)	33.1 (28.6)	32.6 (29.3)	32.3 (29.8)	32.1 (30.2)	31.9 (30.4)	31.8 (30.5)	31.8 (30.7)	31.7 (30.8)
bm (kg/ha)	37.2 (24.1)	34.2 (27.4)	33.2 (28.7)	32.7 (29.4)	32.4 (29.9)	32.2 (30.1)	32 (30.5)	31.9 (30.6)	31.8 (30.7)	31.8 (30.9)
lor (m)	15.4 (10.4)	14.2 (11.8)	13.8 (12.3)	13.6 (12.5)	13.5 (12.6)	13.4 (12.7)	13.3 (12.8)	13.3 (12.9)	13.3 (12.9)	13.3 (12.9)

having fewer observations with which to develop model coefficient estimates. The effect of reduced pulse density on RMSE was slight. The behavior shown for ba relative to pulse density in **Figure 2b** is typical of the behavior for the remaining response variables. The median of the simulation distribution of RMSEs does not appear to change as a result of reduction in pulse density until pulse density is reduced to .05 pulses/m^2. The variability of the simulation distributions of RMSE values do, however, increase slightly for reductions in pulse density.

Reduced sample sizes clearly resulted in increased σ_ε values, but they had a detrimental effect on the inferences that can be made from model RMSEs. For models fitted with 105 observations there was little difference between σ_ε and RMSE values. However, as sample size decreased the fit of the model was purported to improve according to RMSE, but in reality declined with respect to fitting the entire population according to σ_ε values. This phenomenon can be seen in **Table 2** from the increasing deviations between σ_ε and RMSEs to a maximum of over 50% for 15 observations for vol and bm.

Mean estimator

The behavior of the regression estimator of the mean relative to pulse density and sample size, shown in **Figure 3** for lor, was similar to that observed for the remaining response variables. The sampling distributions of mean residuals indicate that the mean estimator is nearly unbiased and that, as we would expect, the variability of mean residuals increases for smaller sample sizes (**Figure 3b**). However, the precision is fairly constant from the largest sample size down to a sample size of 45 observations. Below this threshold the widths of 95% CIs increase more rapidly as sample size decreases. Reduction in pulse density (**Figure 3b**) does not appear to affect the bias or variance of the regression mean estimator.

SE performance

SEs varied methodically from a negative to a positive bias as sample sizes increased from 15 to 105 plots (**Figure 4b**). The trends in SE behavior displayed in **Figure 4** when used to estimate bm is very similar to the trend observed when the estimator was used for the other variables. SEs were positively biased for 55 or more observations, approximately unbiased for 45 observations, and gradually increasingly negatively biased for fewer than 45 observations. No effect from pulse density was detected on the SEs (**Figure 4a**).

The effect of variance estimator bias on confidence intervals as a result of sample size was more dramatic. There was a sharp decline in the percentage of 95% asymptotic confidence intervals that cover the population mean following reductions in sample size to fewer than 35 observations (**Table 3**). This is perhaps not surprising as asymptotic properties are dependent upon having a reasonably large sample size, although the application will affect which sample size is sufficiently large. Coverage probabilities were nearly unchanged for different pulse densities (not shown).

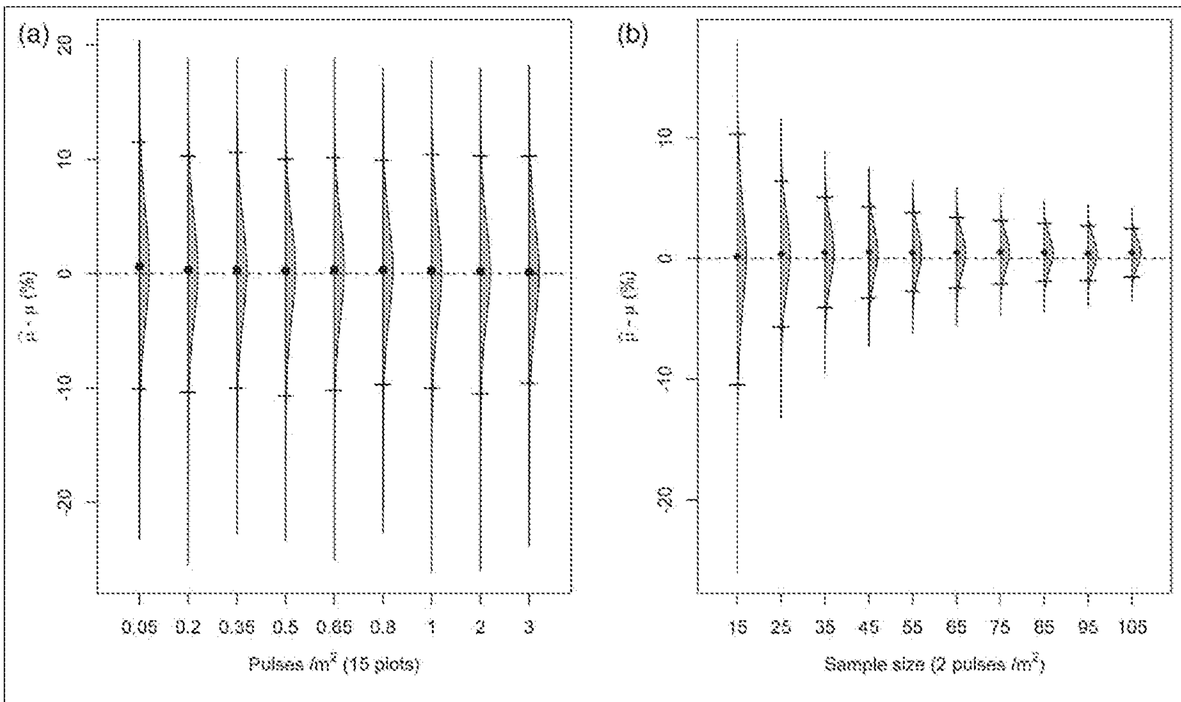

Figure 3. Simulation distributions of differences between estimates of mean lor and the population mean for different (a) pulse densities and (b) sample sizes. Black dots indicate the medians of differences and the horizontal lines indicate bounds of 95% empirical confidence intervals.

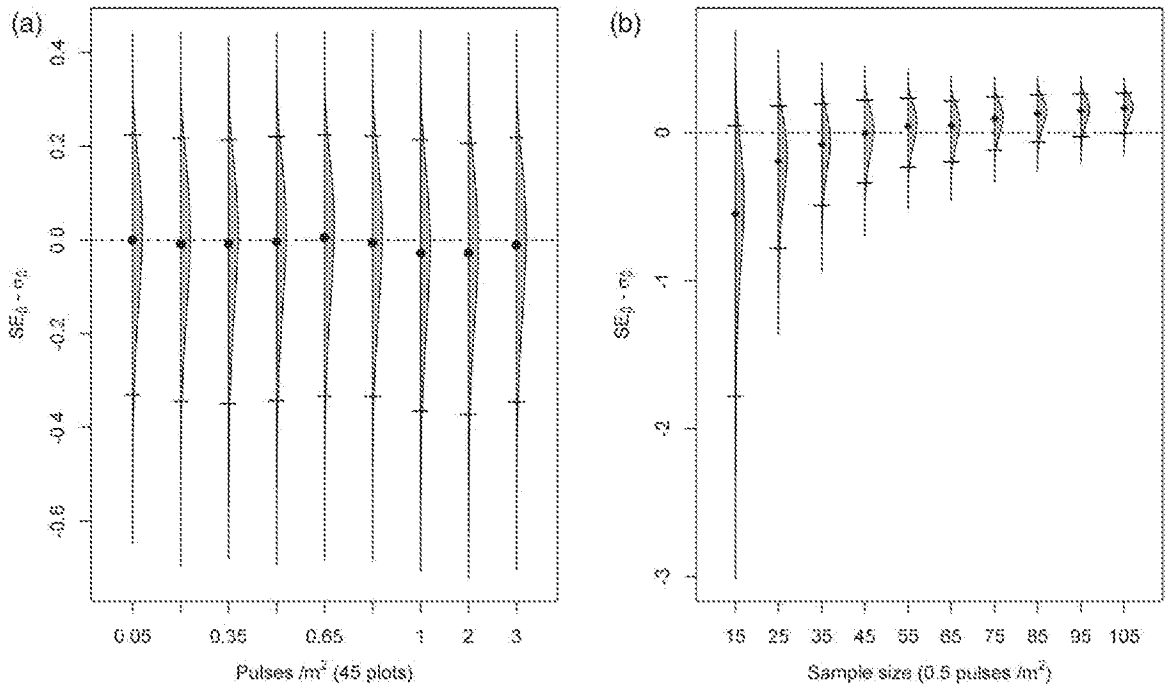

Figure 4. Simulation distributions of percent differences between sample SEs and $\sigma_{\hat{\mu}}$ values for bm by (a) pulse density and (b) sample size. Black dots indicate the medians of differences and the horizontal lines indicate bounds of 95% empirical confidence intervals.

Relative precision

Relative precision varied for a given sample size amongst the response variables (**Table 4**), but the precision was not affected by pulse density (not shown). Our primary interest in relative precision was the region where relative precision is equal to or smaller than 1.0. This is the region where the regression estimator provides greater or equal estimation precision than with 120 plots alone. Each response variable was sufficiently correlated with lidar metrics to yield some gain in precision from using the regression estimator. The threshold at which design effect was less than 1.0 varied from 35 to 75 plots. For sample sizes of 75 observations or more, regression estimation was more precise than plot-only estimation on average for the forest inventory variable stems, the variable that fared poorest. The results for pulse density were not reported because they do not appear to have any effect on relative precision for the densities examined.

Table 3. Coverage probabilities by sample size for 1 pulse/m^2 for analytical 95% confidence intervals.

	Coverage probability				
Sample size	ba	stems	vol	bm	lor
15	0.80	0.79	0.79	0.79	0.81
25	0.88	0.87	0.88	0.87	0.88
35	0.90	0.91	0.90	0.91	0.90
45	0.92	0.93	0.92	0.91	0.92
55	0.93	0.95	0.92	0.93	0.93
65	0.93	0.95	0.94	0.93	0.94
75	0.94	0.96	0.94	0.95	0.95
85	0.94	0.97	0.95	0.95	0.95
95	0.95	0.97	0.95	0.96	0.96
105	0.95	0.98	0.95	0.96	0.96

Discussion

Pulse density

The effect of reduction in pulse density on estimation and inference in a model-assisted inventory with lidar was assessed in several ways. We looked at the effect of pulse density on model RMSE, on mean estimation precision, and on the precision of the SEs. The reduction in information present in lower density data caused model RMSE to

Table 4. Relative precision for regression estimates by sample size for 1 pulse/m^2.

Sample size	ba	stems	vol	bm	lor
15	2.31	6.51	2.52	2.94	2.75
25	1.19	3.31	1.31	1.49	1.39
35	0.81	2.24	0.88	1.01	0.94
45	0.61	1.70	0.67	0.76	0.71
55	0.49	1.37	0.53	0.61	0.57
65	0.41	1.14	0.45	0.51	0.48
75	0.35	0.98	0.38	0.44	0.41
85	0.31	0.86	0.34	0.38	0.36
95	0.28	0.77	0.30	0.34	0.32
105	0.25	0.69	0.27	0.31	0.29

Note: The denominator sample size for relative precision was fixed at 120.

increase slightly. However, the effect of reduction in pulse density was so small that it did not carry through to impact inference. There was no observed effect of pulse density on mean estimators. Although we demonstrated it empirically, this result can also be inferred because residual variability is for the most part not affected by pulse density. If residual variability is not affected by pulse density then the variability of mean estimates will not be affected by pulse density. Furthermore, any effect of pulse density on the variability between mean estimates will be less than the small observed effect of pulse density on residual variability as the effect is reduced by approximately $\sqrt{n}$ on average, which can be seen in the denominator of our SE formula. We note too that because our analysis included a step for creation of a new DTM from the thinned lidar, our inferences can be applied conservatively to cases where a high-quality DTM is already available.

The empirical study by Parker and Glass (2004) performed with .5 and 1 m lidar pulse spacings corroborates our results in that they did not find an effect of pulse density on estimation precision when using a double sampling estimator, although they looked at a much smaller range of densities. We are not aware of any other studies that detail the performance of model-assisted estimators relative to pulse density. As previously mentioned, there are a number of studies that have looked at model precision relative to pulse density for forest yield variables. In four studies of particular relevance, two of the studies found increased RMSE as a result of lower pulse density (Magnusson et al., 2007; Gobakken and Næsset, 2008), while two others did not (Holmgren, 2004; Maltamo et al., 2006). We found that there was a slight but negligible effect of pulse density on model RMSE. The discrepancy between our findings and other findings may result because there are substantial methodological differences between our study and other studies.

The first methodological difference we noted between this and other studies was that we created highly smoothed DEMs. Smoothed DEMs have less detail, but they appear to be less susceptible to irregularities like pits and spikes. Secondly, our plot size was larger than the plots used in the studies that found a noticeable density effect. For the lowest pulse density we examined there were on average 40 pulses per plot, which is double or more the number of returns for alternate studies that encountered a density effect. The sample statistics used for this study, including percentiles (Serfling, 1980) and ratios, are consistent statistics; this means they asymptotically approach the true population parameter as the sample size increases. It is likely then that sample statistics calculated for our plots with greater area achieved greater stability for a given pulse density than was observed in studies that detected a density effect with smaller area plots.

An important consideration with regards to lidar is that forest inventory is only one of many potential applications of lidar. In some instances lidar may be used for a great number of applications including detailed DTM extraction, individual tree segmentation, canopy surface modeling, and others. For alternative uses a reduced pulse density may have drastic repercussions on what is feasible. For example, it is highly unlikely that accurate individual tree segmentation is feasible with .05 pulses/m^2. However, when forest inventory is the primary concern and the cost of mobilization is not the bulk of the lidar acquisition cost, acquiring reduced pulse density wall-to-wall lidar appears to be a highly viable option to reduce the cost of the acquisition without affecting estimation precision for the variables we examined.

Sample size

As was expected from Equation (2), reductions in sample size had a deleterious effect on the precision of model-assisted estimates. More importantly, reductions in sample size also affected the validity of inferences about model fit and about the sampling distribution of the model-assisted mean estimator. The impact on modeling inferences was demonstrated by looking at model RMSE values relative to σ_ε values. Although models on average performed worse in predicting values for the population for reduced sample sizes, according to σ_ε values, RMSE values would indicate that model performance improves for smaller samples. The same effect was observed for the coefficient of determination (R^2), although we do not present those results here because nearly the same information is captured by RMSE. These results demonstrate that RMSE and R^2 are not very useful criteria to evaluate modeling strategies, especially for small samples, a well-known phenomenon in regression modeling. If, for example, the data were stratified and the models were developed separately by strata the results would appear better on average according to RMSE, but the improvement may in fact be the result of optimistic RMSE values related to having a small sample.

The optimism observed in RMSE values for small samples also resulted in bias in SE values for small samples, which are a function of model RMSE. The biases were small, but coverage probabilities were appreciably biased for small samples. This result indicates that for small sample sizes the SE may provide misleading results. Ideally we prefer unbiased estimators, but even positively biased (estimates of variability are too large on average) estimators of variability can be acceptable, within reason. However, with a negatively biased (too small) variance estimator, the probability that a confidence interval covers the true population parameter will be less than expected. While our results are useful in demonstrating the risk involved in making inference from small samples, we cannot guarantee that the threshold observed here will hold exactly or approximately for other studies. It should, however, serve as a cautionary note and provide further motivation to obtain an adequate sample. In practice it may also prove useful to compare analytical and simulation confidence intervals. This would help assess the bias of analytical variance estimators as well as non-normality and skewness in the resampling distribution of mean estimators.

In some instances users will hope to offset the cost of a lidar acquisition by reducing the number of field plots that are used in forest inventory. To address this question we calculated the relative precision of the regression estimator compared with the plot-only estimator for an SRS design. If we reduce the number of field plots to 75 for use with lidar, this is comparable with the plot-only estimator with 120 plots for the response variable stems. This result is indicated by having a relative precision of 1.0 for this variable for 75 plots. If another of the examined variables is of primary interest, then our estimates indicate that the sample size can be reduced much further to a sample size of 45, although confidence intervals may not be sufficiently reliable for this number of plots.

Limitations

A weakness in this study, common to studies that estimate forest inventory variables, is some variables that were treated as if they were measured, such as volume and biomass, were actually estimated with allometric models. The effect of using predicted biomass and volume on variance estimators is not something that this or other similar studies consider when estimating variability. However, the variables estimated with allometric models are highly correlated with ba. The results for ba can serve in some sense as a reality check on results for bm and vol.

A characteristic of this study that differs from some other studies using lidar to estimate forest attributes is that relatively few heights were measured for each plot. This is the result of using data from an operational forest inventory with a set measurement protocol instead of tailoring a new set of field measurements for this study. This resulted in larger model RMSEs for vol, bm, and lor than we see for ba even though vol, bm, and lor are more closely related to height, the forest characteristic represented by lidar predictors. If we measured each tree height it would likely reduce the residual variability of our model fit and increase the precisions of total and mean estimators for vol, bm, and lor. As a result the standard errors reported here are likely conservative for a given pulse density and sample size relative to what could be obtained using data from inventories in which more tree heights were measured, although this will not introduce bias to our analysis.

Additionally, our field data does not include trees smaller than 8 inches. While this is problematic for management reasons if only merchantable trees are considered, the effect on variance estimates and hence our inference relative to pulse density and sample size is negligible. As the values ba, vol, and bm are exponentially related to height and diameter (which are relatively small for these trees), the magnitude of the contribution of small trees to these variables (and lor, a function of ba and height) is also very slight. If there is interest in representing numbers of small trees, then clearly the effect on stems is very pronounced because small trees are weighted equally with large trees for this response variable.

Conclusions

Lidar is a tool that has gained popularity for remote measurement and monitoring of natural resources. Whether there is benefit to using lidar for forest inventory depends in large part on characteristics specific to each forest. This study should facilitate examination of whether using lidar will provide a gain in precision over an alternative estimation strategy for a specific forest and provide indication of an appropriate sample size and pulse density. Fortunately, almost no effect of pulse density was observed on model precision. We saw equivalent precision with 0.05 pulses/m^2 as with 3 pulses/m^2. In contrast, a sample size effect on residual variability was evident. However, the majority of the influence from sample size on estimation precision resulted from the sampling property that as sample size decreases the variability of mean estimates increases proportional to $\sqrt{n}$. Unfortunately for inference for small samples, we found that central limit theorem based confidence intervals were too narrow for small samples. They also affected how well RMSE values represent model performance for the population. If fewer than 75 observations are used, caution should be taken with respect to inference from confidence intervals and model RMSE values.

Acknowledgements

We would like to acknowledge the United States Forest Services Vegetation Monitoring and Remote Sensing Team from the Pacific Northwest research station for funding and support for data collection and analyses. We would like to thank Jeff Foster from the environmental division of the public works department at the Lewis-McChord Joint Base for providing data and logistical support. Finally, we would like to thank four anonymous reviewers who provided valuable insights and critiques.

References

Andersen, H.E., and Breidenbach, J. 2007. Statistical properties of mean stand biomass estimators in a LIDAR-based double sampling forest survey design. In Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007, Espoo Finland: IAPRS, p. 8–13.

Andersen, H.E., Clarkin, T., Winterberger, K., and Strunk, J.L. 2009. An accuracy assessment of positions obtained using survey- and recreational-grade global positioning system receivers across a range of forest conditions within the Tanana Valley of interior Alaska. *Western Journal of Applied Forestry*, Vol. 24, No. 3, pp. 128–136.

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, No. B4/1; PART 4, pp. 111–118.

Clarkin, T. 2007. Modeling global navigation satellite system positional error under forest canopy based on LIDAR-derived canopy densities. M.Sc. thesis, Univ. of Washington, Seattle, Washington, USA. 99 p.

Corona, P., and Fattorini, L. 2008. Area-based lidar-assisted estimation of forest standing volume. *Canadian Journal of Forest Research*, Vol. 38, No. 11, pp. 2911–2916. doi: 10.1139/X08-122.

Efron, B., and Tibshirani, R. 1993. An introduction to the bootstrap. CRC Press. 456 p.

Gobakken, T., and Næsset, E. 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, Vol. 38, No. 5, pp. 1095–1109. doi: 10.1139/X07-219.

Gregoire, T.G., and Valentine, H.T. 2008. Sampling Strategies for Natural Resources and the Environment. 1st ed. New York: Chapman and Hall/CRC. 474 p.

Holmgren, J. 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, Vol. 19, No. 6, pp. 543–553. doi: 10.1080/02827580410019472.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., and Birdsey, R.A. 2004. Comprehensive database of diameter-based biomass regressions for North American tree species. *General Technical Report NE-319*. Newtown Square, PA: US Department of Agriculture, Forest Service, Northeastern Research Station **45**.

Keyser, C.E. 2010. 49 Pacific Northwest Coast (PN) Variant Overview – Forest Vegetation Simulator. Fort Collins, CO: WO-Forest Management Service Center, USDA-Forest Service. Internal Report.

Kraus, K., and Pfeifer, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 53, No. 4, pp. 193–203. doi: 10.1016/S0924-2716(98)00009-4.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., and Shugart, H.H. 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, Vol. 67, No. 1, pp. 83–98. doi: 10.1016/S0034-4257(98)00071-6.

Lohr, S.L. 1999. Sampling: Design and Analysis. Pacific Grove, CA: Duxbury Press. 494 p.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C., and Culvenor, D.S. 2005. Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, Vol. 214, No. 1–3, pp. 398–412. doi: 10.1016/j.foreco.2004.07.077.

Magnussen, S., Næsset, E., and Gobakken, T. 2010. Reliability of LiDAR derived predictors of forest inventory attributes: a case study with Norway spruce. *Remote Sensing of Environment*, Vol. 114, No. 4, pp. 700–712. doi: 10.1016/j.rse.2009.11.007.

Magnusson, M., Fransson, J.E.S., and Holmgren, J. 2007. Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science*, Vol. 53, No. 6, pp. 619–626.

Maltamo, M., Eerikainen, K., Packalen, P., and Hyyppa, J. 2006. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, Vol. 79, No. 2, pp. 217. doi: 10.1093/forestry/cpl007.

McGaughey, R.J. 2012. FUSION/LDV: Software for LIDAR Data Analysis and Visualization, Version 3.01. USFS.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E., and McKee, W.A. 1999. Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sensing of Environment*, Vol. 67, No. 3, pp. 298–308. doi: 10.1016/S0034-4257(98)00091-1.

Næsset, E. 1997a. Determination of mean tree height of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, Vol. 52, No. 2, pp. 49–56.

Næsset, E. 2004. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*, Vol. 91, No. 2, pp. 243–255. doi: 10.1016/j.rse.2004.03.009.

Næsset, E. 1997b. Estimating timber volume of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry*, Vol. 61, No. 2, pp. 246–253.

Parker, R.C., and Evans, D.L. 2007. Stratified light detection and ranging double-sample forest inventory. *Southern Journal of Applied Forestry*, Vol. 31, No. 2, pp. 66–72.

Parker, R.C., and Glass, P.A. 2004. High-versus low-density LiDAR in a double-sample forest inventory. *Southern Journal of Applied Forestry*, Vol. 28, No. 4, pp. 205–210.

Särndal, C.E., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. New York: Springer-Verlag. 694 p.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–464. doi: 10.1214/aos/1176344136.

Serfling, R.J. 1980. Approximation theorems of mathematical statistics. Wiley. 402 p.

Shen, X., Huang, H.C., and Ye, J. 2004. Inference after model selection. *Journal of the American Statistical Association*, Vol. 99, No. 467, pp. 751–762. doi: 10.1198/016214504000001097.

Sithole, G., and Vosselman, G. 2004. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 59, No. 1, pp. 85–101. doi: 10.1016/j.isprsjprs.2004.05.004.

Strunk, J.L., Reutebuch, S.E., Andersen, H.-E., Gould, P.J., and McGaughey, R.J. 2012. Model-assisted forest yield estimation with light detection and ranging. *Western Journal of Applied Forestry*, Vol. 27, No. 2, pp. 53–59. doi: 10.5849/wjaf.10-043.

Temesgen, H., Monleon, V.J., and Hann, D.W. 2008. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. *Canadian Journal of Forest Research*, Vol. 38, No. 3, pp. 553–565. doi: 10.1139/X07-104.

USFS. 2008. National Volume Estimator Library (NVEL).

Wolter, K.M. 2007. Introduction to variance estimation. Springer Verlag.

Wykoff, W.R., Crookston, N.L., and Stage, A.R. 1982. User's guide to the stand prognosis model, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah, *General Technical Report INT-133*, Vol. 112, pp. 51–52.