Remote Sensing of Environment 173 (2016) 98–108

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass

Timothy G. Gregoire [a,*], Erik Næsset [b], Ronald E. McRoberts [c], Göran Ståhl [d], Hans-Erik Andersen [e], Terje Gobakken [b], Liviu Ene [b], Ross Nelson [f]

[a] School of Forestry and Environmental Studies, 360 Prospect Street, Yale University, New Haven, CT 06511-2104, USA
[b] Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway
[c] U. S. Forest Service, Northern Research Station, St. Paul, MN 55108, USA
[d] Dept. of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden
[e] U. S. Forest Service, Pacific Northwest Research Station, Seattle, WA 98195, USA
[f] Biospheric Sciences Branch, NASA-Goddard Space Flight Center, Greenbelt, MD 20771, USA

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 23 October 2015
Accepted 12 November 2015
Available online 5 December 2015

Keywords:
Sampling
Statistical inference
Variance estimation

ABSTRACT

For many decades remotely sensed data have been used as a source of auxiliary information when conducting regional or national surveys of forest resources. In the past decade, airborne scanning LiDAR (Light Detection and Ranging) has emerged as a promising tool for sample surveys aimed at improving estimation of aboveground forest biomass. This technology is now employed routinely in forest management inventories of some Nordic countries, and there is eager anticipation for its application to assess changes in standing biomass in vast tropical regions of the globe in concert with the UN REDD program to limit C emissions. In the rapidly expanding literature on LiDAR-assisted biomass estimation the assessment of the uncertainty of estimation varies widely, ranging from statistically rigorous to ad hoc. In many instances, too, there appears to be no recognition of different bases of statistical inference which bear importantly on uncertainty estimation. Statistically rigorous assessment of uncertainty for four large LiDAR-assisted surveys is expounded.

1. Introduction

Inventories of forest resources on regional or national scales have evolved considerably over the past century, especially in developed nations of western Europe, North America, Australia and New Zealand. These inventories are applications of sampling methodology, a branch of statistical science wherein the precepts of probabilistic selection, and the consequent inference to population features based on the sample, were beginning to coalesce and become formalized in the early years of the twentieth century.

Advances in forest inventory methods have reflected both the changing paradigms in statistical sampling as well advances in technology. For many decades remotely sensed data have featured prominently as a source of auxiliary information to increase the efficiency of sampling as well as to increase the precision of estimates of aggregate forest resources. Comparatively recently, airborne Light Detection and Ranging (LiDAR) has emerged as a prominent remote sensing technology to enable a three-dimensional image of forest canopy height. Using tools of statistics, this point cloud of laser height measurements can be linked by statistical regression to measurements of forest biomass from a ground sample of forested plots. The resulting model then can be applied to forest areas where field observations of biomass are lacking. LiDAR-assisted forest aboveground biomass (AGB) estimation is thus a prominent example of model-assisted survey sampling, sensu lato, whose precepts in a design-based framework are formalized in Särndal, Swensson, and Wretman (1992).

2. Statistical inference, sampling, and sampling error

One goal of survey sampling is to infer something about the population that has been sampled. Often it is to estimate the aggregate quantity of some characteristic of the population, such as the AGB or C of a forested region. The assessment of sampling error and its expression in the form of a standard error of estimation depends on whether the chosen paradigm of statistical inference is based on an underlying statistical model, or, instead, is based on the sampling design which was employed to acquire the sample data. The differences between these two paradigms of statistical inference have been explicated, inter alia, by Särndal (1978); Gregoire (1998), and McRoberts, Næsset, and Gobakken (2013b). One mode of inference is not necessarily better or worse than the other — their respective advantages and disadvantages simply are different. For either mode of inference, the variance

* Corresponding author.
E-mail address: timothy.gregoire@yale.edu (T.G. Gregoire).

http://dx.doi.org/10.1016/j.rse.2015.11.012
0034-4257/

of an estimator may be derived using universally accepted tools of mathematical statistics. The estimation of this variance, likewise, may proceed according to precepts of model-based or design-based inference. The mode of inference affects the statistical properties of estimators. For example, an estimator of total AGB may be model-unbiased but design-biased, and vice versa. Therefore it is important that the mode of inference be stated explicitly, as in Gregoire et al. (2011), Ståhl et al. (2011) and McRoberts et al. (2013b). Regrettably, much of the literature which reports on LiDAR-assisted sampling and estimation lacks an explicit accounting of the mode of inference, thereby leaving the reader to guess at it, and consequently to speculate on the validity of the derived sampling error.

A probability-based sample is essential for design-based inference, but not for model-based inference. To the extent that a probability-based sample may be approximately balanced in the sense of Royall and Herson (1973), it may be a robust safeguard against a poorly specified model in a model-based framework. An exactly balanced sample is one in which the sample moments of a target or auxiliary variable match the population moments. Model-assisted sampling is a frequently encountered term in the present-day literature on survey sampling generally and forest inventory in particular, and more recently, LiDAR-assisted sampling in the remote sensing literature. The term implies the use of a model to increase the precision of estimation, however variance estimation remains design-based. Following a probability sample from the population of interest (POI), the generalized regression estimator, cf. Särndal et al. (1992), is an asymptotically design-unbiased, model-assisted estimator of the population total.

The variance of an estimator in a design-based framework for inference is the probability-weighted squared deviation of the estimate from its expected value, summed over all samples that may be extracted for a given design for the population of interest. In contrast, the variance of an estimator in the model-based framework is based on the single sample obtained, regardless of the design used to select the sample. Because of this essential difference, spatial correlation among units in the selected sample is an irrelevant concern in a design-based framework (Gregoire, 1998), but may be a relevant concern in a model-based framework (McRoberts et al., 2013b). With model-assisted, generalized regression estimation, the residual variance around the regression line or surface will impact the variance of the estimate of AGB, yet these are two distinct variances. Ultimately we are interested essentially in the latter only, for it is this variance that informs us of how reliable our overall estimate of AGB is. A confusion on this point is evident in the literature on LiDAR-assisted sampling for AGB estimation (Mascaro, Detto, Asner, and Muller-Landau (2011); Saatchi et al. (2011); Asner et al. (2013)), where only the former is assessed. While it is beyond the intended scope of this article to do so, the precepts underlying design-based inference are presented in Särndal et al. (1992), whereas those of model-based inference may be found in Chambers and Clark (2012). Both Chambers (2011) and Magnussen (2015) provide useful overviews and comparisons among inferential paradigms.

Many natural resource inventories, whether LiDAR-assisted or not, rely on a systematic placement of field plots. A design-unbiased estimator of sampling variance following a systematic sampling design does not exist. In its absence, approximations to the variance are commonly used, in the hope that estimators derived from these approximations, while biased, will be sufficiently accurate because the magnitude of bias will be small. Särndal et al. (1992, p. 83) note that these variance estimators are typically conservative (in the parlance of the sampling literature, estimators of the variance of an estimator that exceed the actual variance in expectation are termed conservative (Särndal et al., 1992, p. 83)). Simulation has provided some insight into this matter, cf. Ene et al. (2012) and Ene et al. (2013). Specifically, when using the usual estimator of sampling variance that is design-unbiased following simple random sampling without replacement Ene et al. found this estimator to overestimate the variance following systematic sampling by a substantial amount.

In our applications of LiDAR, to be described, we sometimes post-stratify the cells that are used to tessellate the airborne laser scanning (ALS) flightlines into smaller areal units based on thematic information provided by Landsat imagery. Post-stratification is necessary when separate estimates of AGB for a) different land cover classes, or b) different political or ownership classes, are desired. In the statistical literature, estimation for a subset of the population is alternately referred to as "domain estimation". The estimation of post-strata totals has a direct and deducible impact on the variance and reliability of estimates of post-strata totals, as we show in the Appendix A as Supplementary material for Hedmark County. Inasmuch as variance is inversely related to the size of the sample, estimates for a post-stratum total may be quite imprecise when there are few sampled elements in the post-stratum. In this situation, the accuracy of the estimate of standard error may be quite poor, also.

Because biomass comprises nearly 50% C, the imperative of climate change to restrain the emission of C into the atmosphere has focused considerable attention on the stock of C in the world's forests in the form of tree biomass. Indeed, the United Nations REDD program (UN REDD) (http://www.un-redd.org) was initiated to offer financial incentives for sustainable management of forest resources and its C reserves (McRoberts, Tomppo, Vibrans, & de Freitas, 2013a). In this article we report on the design and implementation of four large surveys which used LiDAR to aid in forest biomass estimation: two in southeastern Norway, one in interior Alaska, and one in Tanzania. All but one are two-tiered surveys but differing with regard to stratification: either pre-sampling stratification of the landbase comprising the POI (Tanzania), post-stratification of the sample a posteriori (Hedmark County), or no stratification. In common with many contemporary, comprehensive inventories of forest resources, the LiDAR-assisted surveys are necessarily complex, thereby complicating their statistical analyses in ways that we make clear. Because of the important role that LiDAR may have in the estimation of tropical forest biomass, the results of such studies must withstand considerable statistical scrutiny. Inappropriately devised and overly optimistic estimates of precision may give rise to unrealistic hopes for the REDD Monitoring, Reporting and Verification (MRV) process by LiDAR surveys.

We pay particular attention to statistical issues that arise in estimation, because credible scientific evidence to support REDD mechanisms must rely crucially on credible statistical procedures. The relevance of this work more broadly is that applications of LiDAR-assisted sampling for REDD + MRV in the tropics is likely to involve complicated sampling designs, as well.

3. Variance estimation for a single-tier design

To set the stage for variance estimation following more complicated LiDAR-assisted sampling designs, we first present a simpler design in order to provide a framework for variance estimation in the two-tier designs.

As reported by McRoberts et al. (2013b), airborne LiDAR data were acquired in the municipalities of Åmot and Stor-Elvdal, Norway, during the summer months of 2006. Field-based estimates of AGB were derived from a network of 250 m^2 field plots that spanned the 1259 km^2 region on a 3 km square grid. These were part of the Norwegian National Forest Inventory (NFI). Biomass estimates only from the $n = 145$ NFI plots that were measured in the years proximate to the LiDAR campaign were used in this study. For subsequent reference, these n plots represent the field sample which we denote by S. The entire study region was tessellated by square 250 m^2 cells. A nonlinear regression model — namely, $Y = f(X;\beta) + \varepsilon$ — was fitted to correlate metrics, X, of the distribution of LiDAR measurements on each grid cell to its AGB, y. The model was fitted to the field-based AGB estimates and LiDAR metrics from the 145 NFI plots.

3.1. Model-assisted estimator of AGB and design-based inference

Using N to denote the number of tessellating grid cells in the study region, the fitted regression was then used to generate a prediction of AGB, say $y_k, k = 1,\ldots,N$, for each 250 m^2 square cell in the study area. Denoting the total AGB in the study region by B, the generalized regression estimator of B, following a simple random sample without replacement (SRSwoR) resulting in the sample S, is

$$\widehat{B} = \sum_{k=1}^{N} \widehat{y}_k + \frac{N}{n}\sum_{k\in S} e_k, \tag{1}$$

as presented in Särndal et al. (1992), Eq. (6.5.3). In Eq. (1), $e_k = y_k - \widehat{y}_k$, where $\widehat{y}_k = f(x_k'\beta)$ is the estimate of y_k provided by the fitted model. A design-unbiased estimator of the variance of $\widehat{B}$ is

$$\widehat{V}\left(\widehat{B}\right) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{k\in S} \frac{\left(y_k - \widehat{y}_k\right)^2}{n-1}, \tag{2}$$

as seen in Särndal et al. (1992, p. 236). W clarify that there is a notational difference in our use of $\widehat{B}$ to denote an estimate of AGB and that of Särndal et al. (1992) who use $\widehat{B}$ to denote an estimator of the coefficient vector of the model. Also, the generalized regression estimator (GREG) is associated with a linear model in Särndal et al. (1992), yet this hardly is a necessary condition to use the generalized regression estimator. As Särndal (2011, p. 360), wrote: "Much recent research has been devoted to the GREG form (2.5) for nonlinear relationships between y and x, … Among these contributions one can mention logistic regression, nonparametric regression, local polynomial regression, splines and other techniques."

3.2. Model-based estimator of AGB and inference

In many applications of statistics outside the realm of sampling, inference is based on the stochastic behavior of the response variable as stipulated by an assumed model for that behavior. Model-based inference in a sampling context for AGB likewise regards the observed AGB on a field plot to be a realization of a random variable whose mean and variance are stipulated by the model, and the model itself provides the basis for determining an estimator of B as well as the variance of this estimator. Slightly paraphrasing McRoberts et al. (2013b): "The assumptions underlying model-based inference differ considerably from the assumptions underlying design-based inference. First, the observation for a population unit is a random variable whose value is considered a realization from a distribution of possible values, rather than a fixed value as is the case for design-based inference. Second, the basis for a model-based inference is the model, not the probabilistic nature of the sample as is the case for design-based inference. Randomization for model-based inference enters through the random realizations from the distributions for population units, whereas randomization for design-based inference enters through the random selection of population units into the sample."

Two features of model-based inference permit it to be applied in instances for which design-based inference is limited or even impossible. First, because model-based inference does not rely on a probability sample, it can be used with data acquired using a much greater variety of sampling schemes. Second, model-based inference can also use data external to the area of interest, meaning it can be used for small areas for which samples sizes may be insufficient for design-based inference and for remote and inaccessible areas for which any kind of sampling is impossible.

This is in distinct contrast to regarding the AGB of a field plot as a fixed, but possibly unknown value, and then relying on the probabilistic selection of the sample to determine the variance of an estimator over all possible samples that might have been selected.

A model-based approach to estimation, as in McRoberts and Westfall (2014), takes

$$\widehat{B} = \sum_{k=1}^{N} \widehat{y}_k \tag{3}$$

as the estimator of B. Based solely on the model and conditioning on the sample that was obtained, the model-based estimator of the variance of B is

$$\widehat{V}\left(\widehat{B}\right) = \sum_{k=1}^{N}\sum_{j=1}^{N} Z_k'\widehat{V}\left(\widehat{\beta}\right)Z_j, \tag{4}$$

where $\widehat{V}(\widehat{\beta})$ is the estimated covariance matrix of the $\widehat{\beta}$ estimates, and Z_k is the vector of partial derivatives, $\partial f(x_k'\beta)/\partial\beta$, evaluated at $\widehat{\beta}$. This estimator of variance is based on a series expansion, and should be regarded as an approximation.

4. Study design in Hedmark County, Norway

4.1. Hedmark County

This study was also conducted in Norway, but was considerably more complex, which vastly complicates the estimation of sampling error and the precision of the results that were obtained. Hedmark County (HC) encompasses approximately 27,390 km^2 in southeastern Norway bordering Sweden. Much of the County is heavily forested and mountainous: elevations range from 119 to 2178 m a.s.l. As elevation rises from south to north, forest productivity generally decreases. Norway spruce (*Picea abies* (L.) Karst) and Scots pine (*Pinus sylvestris* L.) are the dominant conifers, and Downy birch (*Betula pubescens* Ehrh.) is the dominant broadleaf species.

By using official land use maps coupled with Landsat TM data, we were able to assign each area cell within HC to one of eight non-overlapping land cover classes. There were four forest productivity cover classes (1) high (8%), (2) medium (13%), (3) low (16%), and (4) young forest (17%), which collectively cover 51% of the land area of HC. The remaining four classes are (5) non-productive forest (11%), (6) mountain (28%), (7) open water (5%), and (8) developed areas (5%). An objective of this LiDAR-assisted survey was to estimate biomass per hectare for each of these cover classes but the last, and to provide a credible estimate of uncertainty for each.

4.2. Norwegian National Forest Inventory plots

The Norwegian NFI plots are located on a 3 km square grid in the forested, non-mountainous portion of the county. Above the coniferous tree line in the mountain areas only plots on a 3 × 9 km grid are inventoried by the NFI. In the parlance of statistical sampling, these plots are a systematic sample of HC land area. Each plot is circular and covers a ground area of 250 m^2. The diameters of all trees on the plot greater than 5 cm were measured. The total aboveground biomass (AGB) of the plot was estimated as the sum of the AGB of the measured trees. Tree AGB was predicted from species-specific allometric models that had previously been fitted by Marklund (1988).

4.3. LiDAR sampling in HC

Airborne scanning LiDAR data were acquired from July to September 2006. For scanning LiDAR data acquisition, 53 parallel E–W flight lines, spaced at 6 km intervals, were established. Each was centered over a series of NFI plots. In total, the flight lines overflew 705 NFI plots that had been measured in 2005–2007 (see Fig. 1). Collectively the scanning LiDAR flight lines spanned 4570 km and covered an area of 2297 km^2, or 8.4% of HC.

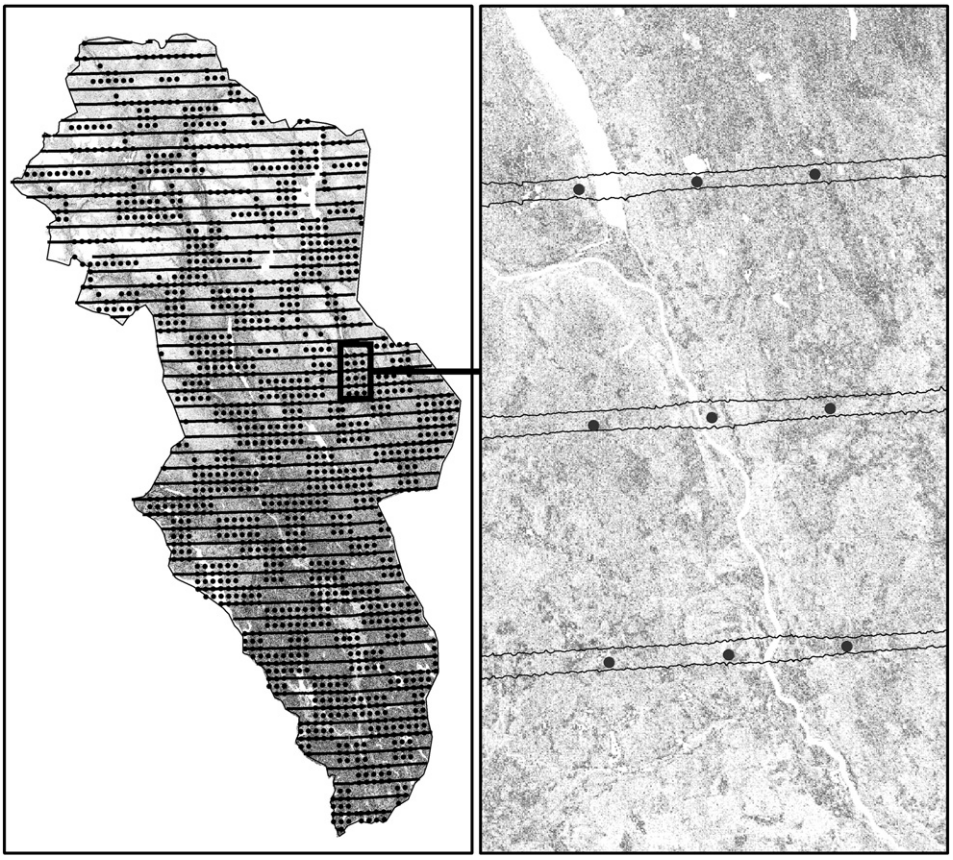

Fig. 1. NFI plot locations in Hedmark County. Gray-shaded background indicates above-ground biomass density (dark is high biomass; light is low biomass). The 53 horizontal lines are the airborne laser scanning flight lines (spacing of 6 km). Only those plots measured in 2005–2007 were used in this study, hence the appearance of gaps in plot coverage.

Average flight altitude was approximately 800 m a.g.l. at a flight speed of 75 ms^{-1}. A maximum half scan-angle of 17° resulted in a swath width of approximately 500 m, therefore, laser height measurements were collected for 250 m on both sides of the flight line.

The laser scanning data along each 500 m swath were partitioned into 250m^2 square cells which matched the ground area of each NFI forest plot. Typically, each cell gives rise to hundreds of laser height measurements, thereby constituting a distribution of height measurements. Summary measures of each distribution were computed, such as the height quantiles, coefficient of variation, mean height. The density decadal quantiles were also computed as the proportion of echoes with heights greater than 0%, 10% … 90% between 1.3 m above ground and the 95th percentile.

A regression model was fitted to link the biomass of each NFI plot to the summary measures of the height distribution that was computed for the plot, that is, one or more of the summary measures of the height distribution were used as predictors of biomass. Generally speaking, the predictor variables that were selected varied among different cover classes, because the optimal set of predictors for, say, the low productivity forest class might be different from those that are optimal to predict biomass for the high productivity forest class. No regression was fitted for the water cover class. When biomass was found on an NFI plot in the water class, the regression that had been fitted for the medium site productivity class was used.

4.4. Sampling design

With this arrangement of NFI ground plot measurements and laser measurements, a two-level systematic sampling design is presented. Each 500 m E–W swath is a primary sampling unit (PSU) whereas each 250 m^2 cell is a secondary sampling unit (SSU). The flight lines actually flown constitute the sample of PSUs, and the NFI plots that were overflown constitute the sample of SSUs.

Whereas HC was stratified by cover class for purposes of computing the land area within HC in each land cover class, it was not feasible to conduct the LiDAR sampling separately in each of these population strata. It was feasible, however, to group cells of the flight lines that were in the same cover class into their appropriate cover class. Post-stratification is a widely used tool in survey sampling, cf. Smith (1991) and McRoberts, Gobakken, and Næsset (2012). A distinguishing, and complicating, feature of the HC design is that post-strata estimates of biomass are correlated owing to the occurrence of multiple cover classes on the same flight line.

Gregoire et al. (2011) considered the HC survey as an application of a two-stage sampling design, and proceeded to use a model-assisted, sensu lato Särndal et al. (1992), two-stage regression estimator with the sample design as the basis for inference.

Aside from the inferential basis, another salient difference is the estimation of the standard errors of estimates for HC as a whole and for each of the post-strata cover classes. As aptly observed by McRoberts et al. (2013b), for complex sampling designs such as this one the estimation of variance can be a challenging task. Despite the challenge, we firmly advocate that variance ought not be estimated in an ad hoc fashion, but rather must be firmly legitimized on the basis of statistical principles.

In the sequel, we report methods of estimation and inference based on consideration of the HC LiDAR survey as a two-stage sampling design.

4.5. LiDAR-assisted estimation of biomass in HC

Indexing sampled PSU's by "i", the cells comprising PSU_i by "k", and denoting biomass on PSU_i by B_i, a generalized regression estimator of B_i is

$$\widehat{B}_i = \sum_{k=1}^{N_i} \widehat{y}_{ik} + \frac{N_i}{n_i} \sum_{k=1}^{n_i} e_{ik} \tag{5}$$

as given by Särndal et al. (1992, Eq. (8.9.6)). In Eq. (5), n_i denotes the number of NFI plots in PSU_i. In Eq. (5), $\widehat{y}_{ik}$ is the predicted biomass of the kth SSU from the LiDAR-based regression and e_{ik} is the regression residual. The first term in Eq. (5) is commonly referred to as a synthetic estimator; as indicated, it is based on the sum over all the N_i SSUs within PSU_i. The second term is a Horvitz–Thompson-like adjustment to the synthetic estimator, an adjustment that makes $\widehat{B}_i$ aysmptotically design-unbiased. The adjustment term is based on only those SSUs that coincide with the n_i NFI plots in the second-stage sample from PSU_i.

The analogous estimator of biomass per hectare for PSU_i for the cth cover class is based only on the SSUs that are in that cover class:

$$\widehat{B}_{ic} = \sum_{N_{ic}} \widehat{y}_{ik} + \frac{N_i}{n_i} \sum_{n_{ic}} e_{ik}. \tag{6}$$

Using m to indicate the number of PSUs in the sample, biomass for HC overall is the probability weighted combination of all the $m = 53$ PSU estimates, namely

$$\widehat{B} = I \sum_{m} \widehat{B}_i, \tag{7}$$

where $I = 12$ is the first-stage systematic sampling interval between successive PSU flight lines. For the county-level estimate of biomass for cover class c:

$$\widehat{B}_c = I \sum_{m} \widehat{B}_{ic}. \tag{8}$$

The variance of $\widehat{B}$, denoted by $V(\widehat{B})$, arises from the variability in biomass among the SSUs within each flight line, as well as variability among the M PSUs in HC. It may be very generally expressed as

$$V\left(\widehat{B}\right) = V_I\left(\widehat{B}\right) + V_{II}\left(\widehat{B}\right), \tag{9}$$

where the first term on the right accounts for the variance in $\widehat{B}$ that is attributable to variation of AGB among all PSUs in HC, and the second term accounts for variance in the estimated biomass that is imparted due to variation in biomass among SSUs within each PSU. Whereas $V(\widehat{B})$ is reckoned over all PSUs and all SSUs within PSUs in HC, the estimator of that variance is necessarily based only on the data collected in the realized sample. In their result 8.9.2, Särndal et al. (1992) derived an expression for the approximate variance of $\widehat{B}$ for a general probability, nonsystematic sampling design.

An expression analogous to Eq. (9) holds for $\widehat{B}_c$. Details on these variance approximations appear in the Appendix A.

Särndal et al. (1992) also derived a design-unbiased estimator of this approximate variance, which was used in Eqs. (18)–(27) of Gregoire et al. (2011). It is standard practice to use an estimator of variance that is unbiased under a simple random sampling design, with the expectation that it will very likely yield an overestimate of the actual sampling variance from a systematically selected sample.

We emphasize that the variance of an estimator is necessarily nonnegative, however a sample-based estimator of the variance need not be. In particular, for the estimator of variance derived in Särndal et al. (1992) the very conditions imposed to achieve design-unbiased estimation of the variance somewhat perversely also allow the estimator to be negative, especially when dealing with small samples. Here the overall size of the HC sample is large, however the subset of the overall sample in some of the less extensive cover classes resulted in small samples for those classes.

4.6. Results for Hedmark County

The results for two-stage model-assisted estimation of biomass, expressed on a per hectare basis by cover class, are shown in Table 1. Comparison of uncertainty estimates under simple random sampling and sophisticated two-stage designs is complicated by large differences in the designs and assumptions. To have comparable designs, we compared AGB estimates based on the field survey only assuming two-stage sampling (the NFI plots being grouped in clusters) against corresponding estimates assuming two-stage sampling with the LiDAR. These results appear in Gobakken et al. (2012) and Næsset et al. (2013), and differ slightly from those presented in Gregoire et al. (2011).

For all cover classes where the estimated standard error was positive, the SE for the LiDAR-assisted estimate was less than the corresponding SE for the estimate based on the field plots alone. In percentage terms, sometimes the LiDAR-assisted SE was less than half the SE for the field plot estimate. For Young Forest class, the percentage SE of the LiDAR-assisted estimate was greater. The reduction in SE is one measure of the gain from using LiDAR data as a source of auxiliary information. For all cover classes combined, the LiDAR-assisted SE was about 2/3 of the SE from field plots alone.

Simulations by Ene et al. (2012) and Ene et al. (2013) subsequently verified that variance estimators that are design-unbiased under a simple random sampling design tend to be very conservative in a systematic framework, that is, they may overestimate the actual sampling variance considerably. Ene et al. (2012) found that an alternative variance estimator based on successive differences was far less biased.

No explicit account has been made for variation in $\widehat{B}$ that is imparted by the possible lack of fit of the allometric model for individual tree biomass that was used to derive an estimate of AGB for each NFI field plot. From a statistical viewpoint, the estimate of AGB is made conditionally on the predictions of biomass given by the fitted allometric models of Marklund (1988). The asymptotic design-unbiasedness of the generalized regression estimator therefore holds with reference to the predicted biomass of HC from these fitted models, and not to the actual AGB of

Table 1
Model-assisted biomass estimates (Mg ha $^{-1}$) by cover class for Hedmark County.

	Area (%)	n[a]	NFI alone		LiDAR assisted	
			Mean	SE (%)	Mean	SE (%)[b]
Productive forest:						
High	5	48	98.5	17.6 (17.9)	120.0	11.1 (9.3)
Medium	13	105	90.6	11.8 (13.0)	90.6	4.8 (5.3)
Low	16	141	49.0	5.7 (11.6)	39.8	5.6 (14.1)
Young	17	151	33.0	4.7 (14.2)	40.4	NA[c]
All productive forests	51	445	64.4	5.6 (8.7)	60.7	4.5 (7.4)
Nonproductive forest & nonforest:						
Nonproductive forests	11	83	19.6	3.5 (17.9)	26.9	NA
Mountain areas	28	95	18.5	5.2 (28.1)	5.1	0.9 (17.6)
Water	5	36	0.0	0.0 (0.0)[d]	2.5	0.0 (0.0)
All nonproductive & nonforest	44	214	15.3	2.7 (17.6)	10.2	NA
All	95	659	46.1	3.0 (6.5)	38.1	1.9 (5.0)

[a] Number of NFI field plots.
[b] SE is the estimated standard error, in units of biomass and as a percentage of the estimated mean.
[c] NA denotes the occurrence of an estimate of SE < 0.
[d] For the Water cover class, SE = 0.0 because aboveground biomass was uniformly absent.

HC. For those who are versed in forest inventory, this is a quite familiar situation when estimating wood volume of a stand or forest or region. We lack statistical tools to account for model-based error in the allometric models. Indeed it is not at all clear whether lack of fit of the allometric models adds bias to the LiDAR-assisted generalized regression estimator presented here, or whether allometric-model lack of fit affects only the variance of $\widehat{B}$. It seems virtually certain, however, that simply adding a term to the design-based or model-based variance of $\widehat{B}$ in an attempt to account for the effect of uncertain allometry is arbitrary and not easily justified from a statistical standpoint.

4.7. Model-based inference for HC

In contrast to the design-based inferential approach of Gregoire et al. (2011); Ståhl et al. (2011) adopted a model-based inferential approach. "The objective of this study was to develop and apply a general framework for model-based estimation and error assessment, accounting for both sampling and model errors, in cases where regression models are applied to predict the target variables." For sake of comparison, we present this model-based approach to estimate the AGB of HC from the LiDAR-assisted sample of HC. The authors presumed a statistical model, possibly nonlinear in the parameters α_h, $B_h(x) = g(x, \alpha_h, \in)$, to relate AGB to one or more LiDAR metrics, x, in the hth post-stratum. In expectation, $E[B_h(x)] = g(x, \alpha_h)$. Letting M denote the total number of flight lines spanning HC and $E[B_h]$ to be the total expected (under the model) AGB of HC in the hth post-stratum, the authors assumed that

$$E[B_h] = \sum_{i=1}^{M} E[B_{ih}] \tag{10}$$

where

$$E[B_{ih}] = \sum_{t=1}^{T_{ih}} g(x_{iht}, \alpha_h), \tag{11}$$

where T_{ih} are the number of cells in stratum h in the ith flight line. Therefore, the total expected AGB of HC is presumed to be

$$E[B] = \sum_{i=1}^{M} \sum_{h=1}^{H} E[B_{ih}]. \tag{12}$$

It is this model-based total that is the population attribute to be predicted.

The LiDAR data from the NFI plots that were overflown on m flight lines, coupled with the biomass from those plots were used to estimate α_h for each stratum. Using $\hat{\alpha}_h$ to denote the estimated parameter vector for stratum h, an estimator of $E[B_{ih}]$ is provided by

$$E[B_{ih}] = \sum_{t=1}^{T_{ih}} g(x_{iht}, \hat{\alpha}_h). \tag{13}$$

To estimate the stratum total AGB, a ratio estimator was used, namely

$$\widehat{E[B_h]} = \left(\frac{\sum_{i=1}^{m} \widehat{E[B_{ih}]}}{\sum_{i=1}^{m} N_{ih}} \right) A_h, \tag{14}$$

where N_{ih} is the number of cells in stratum h within the ith PSU, and A_h is the total land area in stratum h in HC. Using the strata estimates from Eq. (14) leads to

$$E[B] = \sum_{h=1}^{H} E[B_h] \tag{15}$$

as the estimated expected AGB in HC. The variance of $\widehat{E[B]}$ was decomposed as

$$V(\widehat{E[B]}) = V_s + V_\alpha, \tag{16}$$

where V_s is the variance in $\widehat{E[B]}$ due to sampling, and V_α is variance introduced by statistical uncertainty of the $\hat{\alpha}$ estimates.

Details about an estimator of $V(\widehat{E[B]})$ may be found in Ståhl et al. (2011).

4.8. Model-based results for HC

The county-level estimate of AGB provided was 40.3 Mg ha^{-1}, broken down across the post-strata as shown in Table 2. This estimate exceeds the model-assisted estimate of 38.1 Mg ha^{-1} shown in Table 1 and has an estimated standard error that is approximately two-thirds the magnitude of the model-assisted estimated standard error.

5. Study design in Upper Tanana Valley, Alaska

The region of this study encompassed 2012 km^2 surrounding the towns of Tok and Tanacross of interior Alaska, near the state's eastern border with the Canadian province of Yukon Territory. The forested area of this region encompassed 1639.13 km^2. Its lowland forests host predominantly white spruce (*Picea glauca*) and black spruce (*Picea mariana*), and its upland forests host paper birch (*Betula papyrifera*) and quaking aspen (*Populus tremuloides*).

In contrast to the HC study, the Upper Tanana Valley (UTV) study can be described as a "single-stage cluster sample with a model-based estimate of biomass within each cluster" (Andersen, Strunk, & Temesgen, 2011). Similar to HC, however, in UTV the regression linking the field ascertainment of AGB to LiDAR metrics was based on a subsample of plots within clusters. Owing to concern about the non-probabalistic selection of the field plots used in this study, Andersen et al. (2011) adopted a model-based approach to inference.

5.1. LiDAR sampling in UTV

LiDAR data were acquired during June 2009. The $m = 27$ flight lines varied in orientation, as seen in Fig. 2, because of the mountainous terrain. Flight lines were spaced approximately 2.5 km apart, were 240 m wide, therefore covering 9.6% of the study region shown in Fig. 2. In contrast to the HC study, the UTV field plots were established after the flight lines had been established.

Average flight altitude was approximately 750 m a.g.l. at a flight speed of 82 ms^{-1}.

Similar to the HC study, in UTV the LiDAR data along each 240 m swath were partitioned into 324 (18 × 18) m^2 cells, for which summary measures of each distribution of laser heights, similar to those used in the HC study, were computed. Unlike HC, the cells were not post-stratified into distinct cover classes.

Table 2
Model-based LiDAR-assisted biomass estimates (Mg ha^{-1}) by cover class for Hedmark County.

	LiDAR assisted	
	Mean	SE (%)
Productive forest:		
High	133.8	6.1 (4.5)
Medium	97.8	3.4 (3.5)
Low	47.4	2.2 (4.6)
Young	44.6	3.6 (8.0)
All productive forests	67.7	2.2 (3.2)
Nonproductive forest & nonforest:		
Nonproductive forests	27.4	2.4 (8.7)
Mountain areas	6.0	0.7 (11.5)
Water	3.2	0.3 (9.5)
All	40.3	1.2 (2.9)

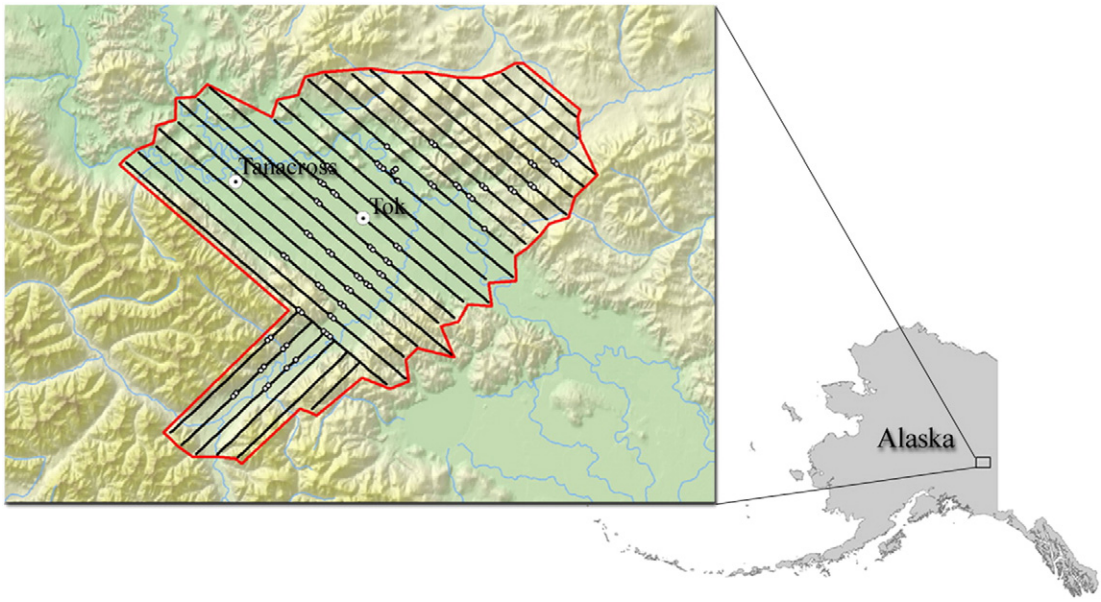

Fig. 2. Upper Tanana Valley study region in Alaska. The solid lines are the airborne laser scanning flight lines (spacing of 2.5 km).

5.2. Field plots for biomass

A total of 79 forested field plots within the LiDAR swaths were established. Each of these circular plots covered an area of 1/30 ha. In view of the difficulty of field data collection, plots were established in pairs, 600 m apart, within 1 km of a road, trail, or river which permitted helicopter access. Within these accessible areas, plots were randomly located to encompass a variety of forest compositions.

Tree diameters and other measures of size were recorded (Andersen et al., 2011). Biomass was predicted for each tree on the plot using the allometric models of Yarie, Kane, and Mack (2007), and then accumulated to derive the predicted aboveground tree biomass on the plot.

A regression model was fitted to link the biomass of each field plot to the summary measures of the height distribution that was co-located with the plot, as in HC.

With a regression fitted to link predicted biomass at the plot-level to LiDAR metrics based on the 79 field plots, this regression was then used to predict biomass on all of the remaining 324 m^2 cells in the LiDAR strips.

5.3. LiDAR-assisted estimation of biomass in UTV

As with the HC study, here too each flight line can be regarded as a PSU and each cell within it as a secondary sampling unit. Because biomass had been predicted for all cells in each flight line, Andersen et al. (2011) used the synthetic estimator of Eq. (5) as the estimator of biomass on the flight line PSU. That is, for the *i*th LiDAR strip,

$$\widehat{B}_i^{'} = \sum_{N_i} \hat{y}_{ik} \tag{17}$$

where, as above, $\hat{y}_{ik}$ is the biomass predicted from the LiDAR-based regression. The necessary nuance of difference between this setup and that of HC is that in the UTV study there were some PSUs with no field plots, and therefore field plots could not be regarded as a second stage of sampling within each PSU. In the parlance of the survey sampling literature, the sampling design was a single-stage cluster sample: the AGB of all N_i elements of PSU_i were predicted by using the LiDAR-assisted prediction of aboveground biomass as a proxy.

The single-stage cluster ratio estimator cf. Särndal et al. (1992, Eq. (8.5.2)) and Cochran (1977, p. 250) was then used to estimate total biomass of the study region:

$$\widehat{B} = N\left(\frac{\sum_m \widehat{B}_i^{'}}{\sum_m N_i}\right), \tag{18}$$

where N is the total number of forested cells in the study region.

Bootstrapping was used to estimate the standard error of Eq. (18). Bootstrapping is a resampling technique introduced by Efron (1979) which has seen widespread application as a nonparametric way to estimate sampling variance following a complex sampling design. In the context of survey-sampling, inferences can be based on a bootstrap population that is generated by resampling in such a way as to mimic the original sampling design. This general approach can be applied in either the model-based or design-based paradigms, and can be used as an alternative to analytical variance formulations in the context of complex multi-level sampling designs (e.g. for exploratory analyses, etc.) Andersen et al. (2011) modified the conventional bootstrap resampling procedure to account for finite population sampling using the without-replacement bootstrap developed by McCarthy and Snowden (1985) and extended by Booth, Butler, and Hall (1994). In addition, Andersen et al. (2011) also examined a modification suggested by Sitter (1997), which entails refitting the model with each bootstrap sample in an effort to account for the additional variability that accompanies the selection of the LiDAR-assisting model (Buckland, Burnham, & Hall, 1997).

5.4. Results for UTV

The estimate of AGB provided by this LiDAR study was 40.4 Mg ha^{-1}. The conventional bootstrap SE was 4.6%. After accounting for model selection, the modified bootstrap SE is increased to 8%.

6. Study design in Liwale District, Tanzania

The study area encompassed a large fraction (15,867 km^2) of the Liwale District (LD) located in southeastern Tanzania (Fig. 3). The dominant vegetation type in LD is the miombo woodlands hosting mainly tree species of the *Caesalpinioideae* family (especially of genera

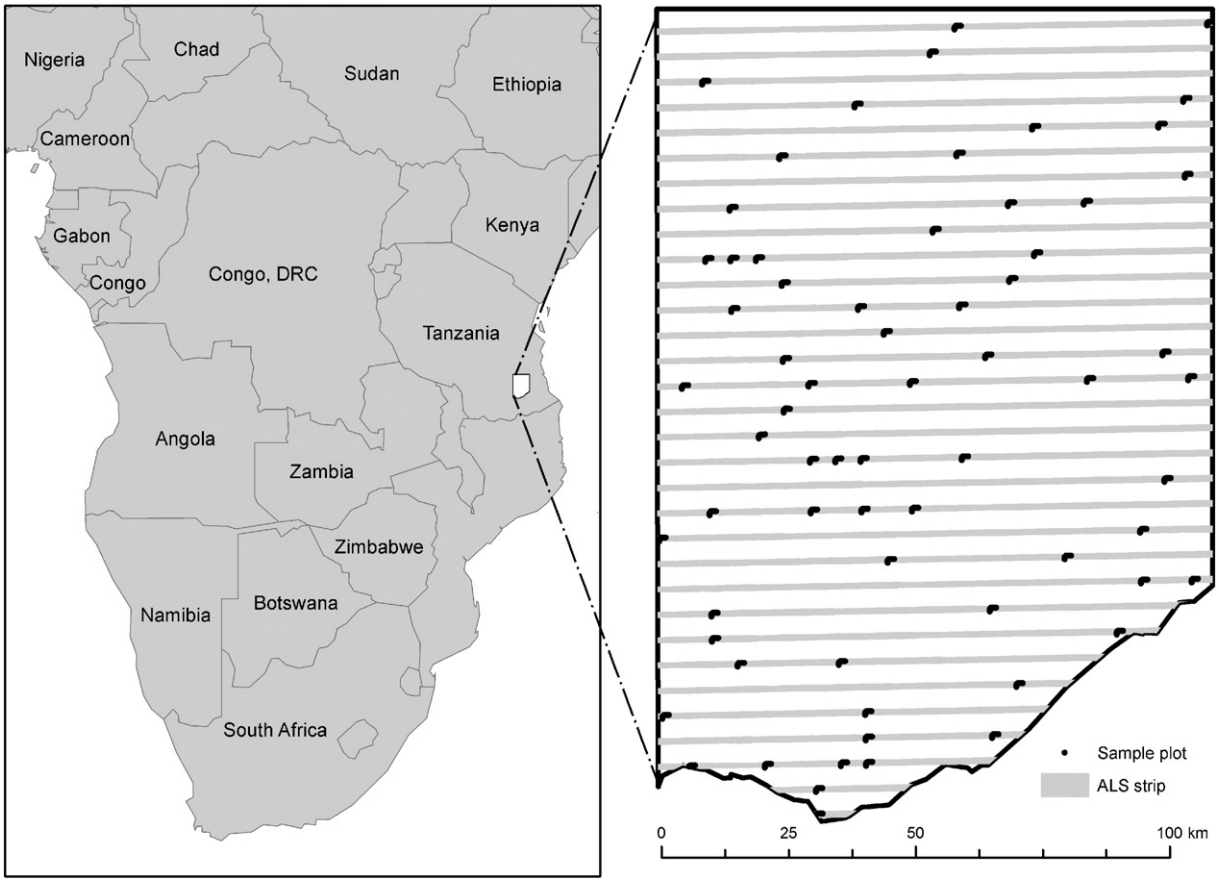

Fig. 3. Liwale administrative district and study area. The solid lines are the airborne laser scanning flight lines (spacing of 5 km) and NAFORMA field plots.

of *Brachystegia* and *Julbernadia*), with some high value commercial timber species such as *Pterocarpus angolensis* (see Mugasha et al. (2013) and references therein).

Similar to the HC study, the LiDAR study reported here utilized ground measurements from the NFI. Unlike HC, LD was stratified prior to field sampling. In LD, clusters of plots served as PSUs, whereas the plots served as the basis for the regression linking AGB to LiDAR metrics.

6.1. Tanzania National Forest Inventory

Tanzania's National Forestry Resources Monitoring and Assessment (NAFORMA) employs a double sampling for stratification design of clusters of field plots, cf. Cochran (1977, pp. 327–335); clusters were located on a 5 km grid and each comprises 10 circular plots spaced 250 m apart in an L-shaped pattern. Each cluster was assigned into one of 18 predefined strata (Tomppo et al., 2014). We note that double sampling for stratification does not form spatially contiguous strata. Instead it allows strata proportions, $W_h, h = 1, \ldots, H$, based on area representation, to be estimated.

Using SRSwoR, a smaller, second-phase sample of clusters was then selected from among the first-phase sample clusters in each stratum. The plots in these second-phase sampling clusters were subsequently measured by the field crews. Details of the field measurements are provided in the Appendix A.

The field plots for this LiDAR-assisted study were established and measured in 2011 under the NAFORMA program, and revisited in February–June 2012 for re-measurement of trees and precise georeferencing. The biomass of stem and branches of sampled trees was predicted using the allometric models of Mugasha et al. (2013). The estimated plot-level AGB ranged from zero (24 ground plots) to 555 Mg ha^{-1} (Ene et al., 2015).

6.2. LiDAR sampling in LD

Scanning LiDAR data were acquired under leaf-on conditions during February–March 2012. The average flight altitude was approximately 1320 m a.g.l., at a ground speed of 77.2 ms^{-1}. Thirty-two parallel flight lines were flown in an east–west direction, spaced 5 km apart, as shown in Fig. 3, covering nearly 26% of the study area. The average strip width was approximately 1350 m, which is shorter than the distance spanned by each leg of the L-shaped clusters (1500 m). Consequently, at most eight plots per cluster in LD were covered by scanning LiDAR data; generally speaking, a variable number of plots in each first-phase cluster were overflown for LiDAR data collection, a factor which further complicates the statistical analysis. From the height distribution of the LiDAR echoes co-located with the 15 m radius plots, LiDAR metrics were derived following the approach described above for HC (Ene et al., 2015).

Over the seven forested strata present in LD there were $m_1 = 626$ first-phase clusters of 10 plots, of which $m_2 = 65$ had been selected for the second phase of field sampling in LD. See Table 3 for sample sizes by stratum in both phases of sampling in LD. When needed in the sequel, S_{1h} shall indicate the first-phase sample of m_{1h} clusters in the hth stratum, and S_{2h} shall indicate the second-phase sample of m_h clusters in the hth stratum.

6.3. LiDAR-assisted estimation of biomass in LD

A regression model was fitted to link the plot-level AGB estimates to the LiDAR metrics. The fitted regression was used to predict biomass on all of the remaining first-phase cluster plots within the scanning LiDAR strips. Assuming SRSwoR at both sampling occasions, an appropriate

Table 3
Distribution of clusters and field plots by NAFORMA strata.

	NAFORMA stratum[a]							
	5	6	7	8	9	11	12	
First stage							Total	
	m_{15}	m_{16}	m_{17}	m_{18}	m_{19}	$m_{1,11}$	$m_{1,12}$	m_1
Number of clusters	160	121	118	67	46	93	21	626
	n_{15}	n_{16}	n_{17}	n_{18}	n_{19}	$n_{1,11}$	$n_{1,12}$	n_1
Number of plots	1268	954	936	533	350	744	166	4951
Second phase							Total	
	m_5	m_6	m_7	m_8	m_9	m_{11}	m_{12}	m
Number of clusters	15	11	12	16	3	3	5	65
	n_5	n_6	n_7	n_8	n_9	n_{11}	n_{12}	n
Number of plots	120	88	94	125	24	24	38	513

[a] According to Tomppo et al. (2014), Table 3.

model-assisted estimator for the mean AGB per hectare for stratum h under double sampling is (Mandallaz, 2008, Eq. (5.1))

$$\widehat{\bar{B}}_h^{reg} = \frac{\sum_{k\in S_{1h}} \widehat{b}_{kh}}{n_{1h}} + \frac{\sum_{k\in S_{2h}} e_{kh}}{n_h} \tag{19}$$

where $\widehat{b}_{kh}$ is the AGB per hectare predicted by the LiDAR-based regression model for the kth cluster in hth stratum, and $e_{kh} = b_{kh} - \widehat{b}_{kh}$ is the plot-level regression residual. In Eq. (19), $\bar{n}_{1h}$ denotes the average number of plots per first-phase cluster in stratum h that were overflown for LiDAR, and $\bar{n}_h$ denotes the average number of plots per second-phase cluster in stratum h that were overflown.

The mean AGB per hectare for LD was estimated as a weighted sum of the stratum estimates:

$$\widehat{\bar{B}}_{DS}^{reg} = \sum_{h\in H} \widehat{W}_h \widehat{\bar{B}}_h^{reg}, \tag{20}$$

where $\widehat{W}_h$ is the estimated weight for the hth stratum.

The approximated variance estimator for the stratum-level estimates $\widehat{\bar{B}}_h^{reg}$ is (Mandallaz, 2008, p. 81):

$$\widehat{v}\left(\widehat{\bar{B}}_h^{reg}\right) = \frac{\widehat{s}_{bh}^2}{\bar{n}_{1h}^2 m_{1h}} + \left(1 - \frac{m_h}{m_{1h}}\right)\frac{\widehat{s}_{eh}^2}{\bar{n}_h^2 m_h} \tag{21}$$

where the terms in Eq. (21) are detailed in Table 3 and in the Appendix A).

The variance of $\widehat{\bar{B}}_{DS}^{reg}$ was estimated based largely on Cochran (1977, Eq. (12.32)):

$$\widehat{v}\left(\widehat{\bar{B}}_{DS}^{reg}\right) = \sum_{h\in H}\left(\frac{m_{1h}}{m_1 - 1}\right)\widehat{W}_h \widehat{v}\left(\widehat{\bar{B}}_h^{reg}\right) + \sum_{h\in H} \widehat{W}_h \frac{\left(\widehat{\bar{B}}_h^{reg} - \widehat{\bar{B}}_{DS}^{reg}\right)^2}{m_1 - 1}. \tag{22}$$

6.4. Results for LD

The LiDAR-assisted AGB estimate was 58.94 Mg ha^{-1} with an estimated standard error of 1.48 Mg ha^{-1} (2.5%). Compared to the estimated standard error of the direct AGB estimator, which were derived solely from the field observations (Ene et al., 2015), the LiDAR-assisted estimation was approximately three times more precise. Due to the systematic arrangement of plots within a cluster, we suspect that even this comparative precision was understated.

7. Discussion

We have summarized four LiDAR-assisted studies principally to emphasize the straightforward but non-trivial task of properly estimating the variance of estimators of AGB based on a complex sampling strategy, that is, one that may be based on two or more tiers of sampling, stratification and post-stratification, and utilize one or more models in the estimation of AGB. While the purposes served by LiDAR-assisted surveys vary greatly, thereby preventing development of a standard sampling design, there are standard statistical tools that can be brought to bear on assessing the variance of proposed estimators, which have not been universally recognized or employed.

When basing inference on the sampling design, the variance of estimators, and therefore the estimation of sampling error, depends crucially on the design. Analogously, when basing inference on a presumed model, the variance of estimators will depend both on the model and how well the model portrays the stochastic behavior of AGB. The variance of an estimator under both the design-based and the model-based framework is well-defined as the statistical expectation of the squared deviation of the estimator from its expected value. It is more nuanced and complicated than merely identifying various sources of statistical error, and then scaling and adding together the perceived variance of these errors.

With this in mind we exhort researchers who report results from LiDAR-assisted AGB studies to explicitly declare the framework of inference and to demonstrate how an estimator's precision was itself estimated within the adopted framework according to the statistical definition of variance. This approach is absent in many articles on LiDAR sampling that have appeared in this journal and many others. For example Asner et al. (2013, p. 12) assert that "Our analysis may also be improved through analytical modeling of errors (i.e., model-based inference) produced by both the LiDAR-to-carbon and national-scale models…" perhaps. Nonetheless, we are left wondering what basis for inference was used in this article if not model-based? Expressions such as

$$\epsilon_{National(otherpixel)} = \sqrt{\epsilon_{LiDAR}^2 + \epsilon_{DecisionTree}^2} \tag{23}$$

provide little clarity as to how this relates to the variance of an estimator of overall AGB, as universally understood in the sampling literature.

We assert that a most urgent problem facing LiDAR-assisted estimation based on systematic sampling is the very large overestimation of estimator variance by assuming a simple random sampling design. The successive differences estimator may be an alternative, less biased, estimator of variance, cf. Ene et al. (2013).

We have noticed oft-expressed concern over the potential bias of individual-tree biomass predictions provided by inappropriate allometric prediction models. This is very similar to the concern that foresters have had for generations with the use of models to predict bole volume of standing timber trees. For that reason, forestry concerns in many regions of the world have made enormous investments in the development of trustworthy volume prediction models. Even with that investment, foresters and forest surveys are able to design-unbiasedly estimate the aggregate predicted bole volume of a POI, yet not the actual aggregate bole volume. At least not with methods that currently are used. Even if a concerted and widespread effort is made to acquire data needed to develop regionally appropriate individual-tree allometric biomass models, we still will not be in a position to unbiasedly estimate the actual aggregate AGB of forest trees. If modern forest inventory provides a useful guide, the hope for those who estimate aggregate AGB is that the estimate of aggregate predicted AGB will be sufficiently close to actual AGB so the difference between the two will be inconsequential. In the ongoing discussions of MRV for REDD +, there has been no recognition of this matter.

As far as the magnitude of the effect on precision that is caused by using predictions rather than measurements of individual tree biomass, two recent studies – McRoberts and Westfall (2014) and Ståhl, Hekkinen, Petersson, Repola, and Holm (2014) – concluded that the variance from this source is dwarfed by the sampling variance of the survey. However, McRoberts and Westfall (2015) observed that the disparity diminished when the sample was post-stratified.

Studies are currently underway to estimate the change in AGB, namely ΔAGB, between two LiDAR-assisted surveys, again with at least a glance towards meeting the needs of MRV for REDD+. For change estimation the bias in biomass change predictions may be reduced, so long as the same model(s) are used on both occasions.

For ΔAGB estimation, new sampling design considerations will arise. As one example, sampling with partial replacement (Ware & Cunia, 1962) configures sample sizes to allow for optimal estimation of current AGB as well as ΔAGB. Yet it further complicates sampling designs that many would regard as already sufficiently complex.

We have noticed a recent emphasis on mapping rather than estimation, with the implication that mapping is necessary for accurate estimation. For example, Chen, Laurin, and Valenti (2015) asserts "However, substantial uncertainty remains in estimating tropical forest C emissions… accurately mapping the spatial distribution of tropical C stock and its dynamics is vital to reduce such uncertainty." We would counter that accurate mapping is not a necessary condition for accurate estimation of regional total stocks of biomass and C, nor does accurate mapping even guarantee precise estimation.

There are three major elements of any sample survey. One is the sampling design, which concerns the selection of relevant data to enable estimation of the population parameters of interest, for example total C or ΔC. Next is the manner in which these parameters indeed will be estimated. As is evident from the four surveys presented in this article, the design surely will inform, but not mandate, the choice of estimator. Finally, there is the mode of statistical inference which will enable a credible assessment of sampling error. This article is intended as a call for researchers to pay much increased attention to these elements.

Appendix A

A.1. Further details for HC

The term $V_I(\widehat{B})$ in Eq. (9) is the variance in estimated AGB among PSUs in HC. With the sampling strategy employed in HC, there are two impediments to estimating $V_I(\widehat{B})$ in a design-unbiased manner. The first is the systematic selection of the PSU flight lines to accord with the systematic spacing of the NFI plots in the E–W direction. The second is the use of the generalized regression estimator, Eq. (5). The design-bias due to generalized regression estimation is usually small if sample sizes are not too small. The design-bias due to approximating the variance following systematic selection with a SRSwoR alternative can be severe, as evidenced in the simulation results of Ene et al. (2012) and Ene et al. (2013).

The following results are based largely on derivations for Case C in Särndal et al. (1992, sec 8.9), which have been evaluated under the presumption of two-stage SRSwoR design.

Under a presumption of SRSwoR selection of m PSUs from M PSUs spanning HC,

$$V_I\left(\widehat{B}\right)\approx M^2\left(\frac{1}{m}-\frac{1}{M}\right)\sigma_I^2, \tag{A1}$$

where $\sigma_I^2=(M\text{-}1)^{-1}\sum_{i=1}^{M}(B_i\text{-}B/M)^2$ is the variance in AGB among the M PSUs.

In similar fashion, $V_{II}(\widehat{B})$ in Eq. (9) is the variance in estimated AGB due to variation in AGB among SSUs within a PSU. Under a presumed SRSwoR selection of SSUs within each PSU,

$$V_{II}\left(\widehat{B}\right)\approx\frac{M}{m}\sum_{i=1}^{M}\ N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\sigma_{IIi}^2, \tag{A2}$$

where

$$\sigma_{IIi}^2=(n_i-1)^{-1}\sum_{k\in PSU_i}\epsilon_{ki}^2, \tag{A3}$$

and ϵ_{ki} is the difference between B_i and the statistical expected value of B_i under the assisting regression model.

Under the presumed two-stage SRSwoR sampling design, an unbiased estimator of $V_I(\widehat{B})$ is

$$\widehat{V}_I\left(\widehat{B}\right)=M^2\left(\frac{1}{m}-\frac{1}{M}\right)\left(s_I^2-m^{-1}\sum_m\ \widehat{V}_i\right), \tag{A4}$$

where

$$s_I^2=(m-1)^{-1}\sum_m\ \left(\widehat{B}_i-\bar{\widehat{B}}_i\right)^2 \tag{A5}$$

and

$$\widehat{V}_i=N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)\sum_{k=1}^{n_i}\ (n_i-1)^{-1}(\hat{y}_{ik}-\bar{y}_i)^2, \tag{A6}$$

where $\bar{y}$ is the average AGB on the NFI plots in the second-stage sample from PSU_i.

An unbiased estimator of $V_{II}(\widehat{B})$ is

$$\widehat{V}_{II}\left(\widehat{B}\right)=\left(\frac{M}{m}\right)^2\sum_{i=1}^{m}\ N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{IIi}^2, \tag{A7}$$

where $s_{IIi}^2=(n_i-1)^{-1}\sum_{k=1}^{n_i}\ (e_{ik}-\bar{e}_i)^2$ and $\bar{e}_i=\frac{1}{n_i}\sum_{k=1}^{n_i}\ e_{ik}$.

A.2. Further details for LD

Field measurements in NAFORMA: Tree diameters greater than 1, 5, 10 and 20 cm were recorded using concentric circular plots with radii of 2, 5, 10 and 15 m, respectively. The height of every 5th tree in the sample was measured and diameter–height models were developed for predicting the heights of the remaining trees. According to Vesa et al. (2011), a tree was defined as a perennial wooded plant at least 1.35 m tall and with a distinct stem capable of reaching 5 m height in situ. Tree species were also recorded. Cacti, palms, bamboos, and shrubs were not recorded as trees.

Definition of terms in Eq. (21):

$$\widehat{s}_{bh}^2=(m_h-1)^{-1}\sum_{k\in S_{2h}}\left(b_{kh}-\widehat{\bar{B}}_h n_{kh}\right)^2, \tag{B1}$$

where n_{kh} denotes the number of plots overflown on the k second-phase cluster, and

$$\widehat{\bar{B}}_h=n_h^{-1}\sum_{k\in S_{2h}}\ b_{kh}. \tag{B2}$$

Also,

$$\widehat{s}_{eh}^2=(m_h-1)^{-1}\sum_{k\in S_{2h}}\ (e_{kh}-\bar{e}_h n_{kh})^2, \tag{B3}$$

where

$$\bar{e}_h = \frac{\sum_{k \in S_{2h}} e_{kh}}{\sum_{k \in S_{2h}} n_{kh}}. \quad \text{(B4)}$$

References

Andersen, H. -E., Strunk, J., & Temesgen, H. (2011). Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the upper Tanana Valley of interior Alaska. *Western Journal of Applied Forestry*, *26*, 157–164.

Asner, G. P., Mascaro, J., Anderson, C., Knapp, D. E., Martin, R. E., Kennedy-Bowdoin, T., ... Bermingham, E. (2013). High-fidelity national carbon mapping for resource management and REDD+. *Carbon Balance and Management*, *8*(1), 1–14.

Booth, J. G., Butler, R. W., & Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, *89*, 1282–1289.

Buckland, S., Burnham, K., & Hall, P. (1997). Model selection: An integral part of inference. *Biometrics*, *53*, 603–618.

Chambers, R. L. (2011). Which sample survey strategy: A review of three diffeent approaches. *Pakistan Journal of Statistics*, *27*(4), 337–357.

Chambers, R. L., & Clark, R. G. (2012). *An introduction to model-based survey sampling with applications*. Oxford: Oxford University Press.

Chen, Q., Laurin, G. V., & Valenti, R. (2015). Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. *Remote Sensing of Environment*, *160*, 134–143.

Cochran, W. G. (1977). *Sampling techniques*. New York: Wiley.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*, 1–26.

Ene, L., Næsset, E., Gobakken, T., Gregoire, T. G., Ståhl, G., & Nelson, R. (2012). Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach. *Remote Sensing of Environment*, *123*, 579–592.

Ene, L., Næsset, E.,, Gobakken, T., Gregoire, T. G., Ståhl, G., & Nelson, R. (2013). A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys. *Remote Sensing of Environment*, *133*, 210–224.

Ene, L. T., Næsset, E., Gobakken, T., Bollandsås, O. -M., Mauya, E. R., Gregoire, T. G., & Ståhl, G. (2015). Large-scale estimation of aboveground biomass in miombo woodlans in Tanzania using airborne laser scanning and national forest inventory data. *Remote Sensing of Environment* (in review).

Gobakken, T., Næsset, E., Nelson, R. F., Bollandås, O. M., Gregoire, T. G., Ståhl, G., ... Astrup, R. (2012). Estimating biomass in Hedmark County, Norway, using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, *123*, 443–456.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, *28*, 1429–1447.

Gregoire, T. G., Ståhl, G., Næsset, E., Gobakken, T., Nelson, R. F., & Holm, S. (2011). Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, *41*, 83–95.

Magnussen, S. (2015). Arguments for a model-dependent inference. *Forestry*, *88*, 317–325.

Mandallaz, D. (2008). *Sampling techniques for forest inventories*. Boca Raton: Chapman & Hall/CRC.

Marklund, L. G. (1988). *Biomass functions for pine, spruce and birch in Sweden (in Swedish)*. Umeå, Sweden: Tech. rept. Swedish University of Agricultural Sciences, Department of Forest Survey.

Mascaro, J., Detto, M., Asner, G. P., & Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment*, *115*, 3770–3774.

McCarthy, P. J., & Snowden, C. B. (1985). The bootstrap and finite population sampling. *Data Evaluation and Methods Research, Series 2 DHHS Publication No. 85-1369*. Hyattsville, MD: U. S. Department of Health and Human Service.

McRoberts, R. E., & Westfall, J. A. (2014). Uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, *60*, 34–42.

McRoberts, R. E., & Westfall, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, *72*.

McRoberts, R. E., Gobakken, T., & Næsset, E. (2012). Post-stratified estimation of forest area and growing stock using lidar-based stratifications. *Remote Sensing of Environment*, *125*, 157–166.

McRoberts, R. E., Tomppo, E. O., Vibrans, A. C., & de Freitas, J. V. (2013a). Design considerations for tropical forest inventories. *Pesquisa Florestal Brasileira*, *33*, 189–202.

McRoberts, R. E., Næsset, E., & Gobakken, T. (2013b). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment*, *128*, 268–275.

Mugasha, W. A., Eid, T., Bollandsäs, O. M., Malimbwi, R. E., Chamshama, S. A. O., Zabahu, E., & Katani, J. Z. (2013). Allometric models for prediction of above- and belowground biomass of trees in the miombo woodlands of Tanzania. *Forest Ecology and Management*, *310*, 87–101.

Næsset, E., Gobakken, T., Bollandsås, O. M., Gregoire, T. G., Nelson, R. N., & Ståhl, G. (2013). Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sensing of Environment*, *130*, 108–120.

Royall, R. M., & Herson, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, *68*, 880–889.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., ... Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899–9904.

Särndal, C. -E. (1978). Design-based and model-based inference in survey sampling. *Scandinavian Journal of Statistics*, *5*, 27–52.

Särndal, C. -E. (2011). Combined inference in survey sampling. *Pakistan Journal of Statistics*, *27*(4), 359–370.

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

Sitter, R. R. (1997). Variance estimation for the regression estimator in two-phase sampling. *Journal of the American Statistical Association*, *92*, 780–787.

Smith, T. M. F. (1991). Post-stratification. *The Statistician*, *40*, 315–323.

Ståhl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., & Nelson, R. F. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, *41*, 96–107.

Ståhl, G., Hekkinen, J., Petersson, H., Repola, J., & Holm, S. (2014). Sample-based estimation of greenhouse gas emissions from forests — A new approach to account for both sampling and model errors. *Forest Science*, *60*, 3–13.

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

Vesa, L., Malimbwi, R. E., Tomppo, E., Zahabu, E., Maliondo, S., Chamuya, N., ... Dalsgaard, S. (2011). National forest resources monitoring and assessment of Tanzania (NAFORMA). *Biophysical survey. Field manual. Tech. rept*. Tanzania: Ministry of Natural Resources & Tourism, Forestry and Beekeeping Division.

Ware, K. D., & Cunia, T. (1962). Continuous forest inventory with partial replacement of samples. *Forest Science Monograph*, *3*(44), 1–40.

Yarie, J., Kane, E., & Mack, M. (2007). *Aboveground biomass equations for the trees of interior Alaska. Tech. rept. Bulletin*, *115*, Agricultural and Forestry Experiment Station, University of Alaska-Fairbanks.