Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway

Erik Næsset, Terje Gobakken, Ole Martin Bollandsås, Timothy G. Gregoire, Ross Nelson, Göran Ståhl

1. Introduction

There is an urgent need for efficient methods to estimate biomass and carbon stocks and changes in such stocks in tropical countries. The United Nations Collaborative Program on Reduced Emissions from Deforestation and Forest Degradation in Developing Countries (UN REDD) was launched with the aim of contributing to the development of capacity for reducing emissions from loss of forest carbon in developing countries. It is understood that REDD mechanisms must be supported by forest assessment programs that can monitor the carbon stocks. Implementation of so-called REDD demonstrations at local scales within countries and even covering entire nations has now started with funding from donor countries. In for example Tanzania, REDD demonstrations involving local assessment of carbon stocks are ongoing while Guyana as a nation...
has already received payments from Norway (Anonymous, 2011a) for the first commitment period (1 October 2009–31 September 2010) for performances on avoided emissions from deforestation and other inter-millennial performance indicators as defined in the bilateral agreement between Guyana and Norway (Anonymous, 2009).

It is expected that remote sensing will play an important role in future monitoring for REDD. Airborne Light Detection and Ranging (LiDAR) has emerged as one of the most promising remote sensing technologies for estimating above-ground tree biomass. Lately, the potential of airborne LiDAR for local (Næsset et al., 2011) and national forest inventory (Næsset, 2005) and for REDD (Deforestation and Forest Degradation in Developing Countries) monitoring has been emphasized (Angelsen, 2008; Gibbs et al., 2007). For monitoring of larger territories like counties, states, provinces, and nations, sampling with LiDAR is a viable option. Sampling with airborne profiling (Nelson et al., 2003, 2004, 2012) and scanning (Andersen et al., 2009; Næsset, 2005; Næsset et al., 2009) LiDARs as well as spaceborne LiDAR (Boudreau et al., 2008; Nelson et al., 2009a,b) has been proposed and demonstrated for territories ranging in size from a few thousand square kilometers (the state of Delaware, USA; Nelson et al., 2003, 2004) to more than a million square kilometers (the province of Quebec, Canada; Boudreau et al., 2008; Nelson et al., 2009a). LiDAR data has been used as part of two stage cluster (e.g. Andersen et al., 2009) or even three phase (e.g. Boudreau et al., 2008) designs.

Asner (2009) recently proposed a REDD monitoring methodology utilizing airborne scanning LiDAR in combination with field data and other remote sensing data and he conducted a demonstration over 4.3 million ha of the Peruvian Amazon (Asner et al., 2010). REDD demonstrations utilizing airborne LiDAR have also been carried out in other countries, e.g., Laos (Gautam et al., 2010) and Nepal (Anonymous, 2011b). Clearly, many of the early LiDAR-supported biomass assessments conducted as REDD demonstrations do not comply with generally accepted principles in forest inventory, at least not as far as appraisal of uncertainties in the biomass and carbon stock estimates is concerned. Further, many studies with LiDAR continue to be conducted in tropical countries with little or no consideration given to how the acquired data will be analyzed.

For estimation methods to be relevant for reporting purposes for tropical countries under a future REDD mechanism and for countries ratifying the Kyoto Protocol to the United Nations Framework Convention on Climate Change trustworthy estimates of changes in carbon stocks with associated and generally accepted estimates of uncertainties should be provided. In cases where estimates and associated uncertainty estimates are based on surveys involving LiDAR, there is evidently an urgent need to come up with reliable frameworks for estimation and inference. It is not trivial to derive statistically sound estimators for complicated designs involving two or even three phases or stages of sampling.

Efforts to develop estimators for two-phase (double sampling) and two-stage (cluster sampling) designs involving LiDAR and ground samples have been ongoing for several years (Næsset et al., 2009). These designs basically consist of a first phase or stage sample of parallel flight-lines flown over the target area with LiDAR data collected along the swathes (scanning LiDAR) or collected as profiles (profiling LiDAR) along the center line. The flight-lines can be spaced many kilometers apart and they do not give a full “wall-to-wall” coverage. Second phase or stage samples of field plots are then collected on the ground along the flight-lines, often according to a systematic layout. Gregoire et al. (2011) and Ståhl et al. (2011) recently presented estimators appropriate for a design by which LiDAR data are collected along such a sample of individual flight-lines over an area of interest and with a sample of field plots collected on the ground along the LiDAR flight-lines. Gregoire et al. (2011) assumed that the LiDAR flight-lines as well as the field plots are random samples from a finite population and derived model-assisted estimators of biomass and corresponding variance estimators, which rest on the principles of probability sampling. Ståhl et al. (2011) derived model-dependent estimators applicable to the same design, but did not assume that the samples were selected according to probabilistic principles. The latter approach is therefore more flexible because the field plots need not come from the same area or they can for example be selected purposefully, which may be an attractive property especially in the tropics where inaccessibility and limited infrastructure may restrict field work in remote areas. However, the work by Ståhl et al. (2011) assumed that the regression model that relates biomass observed on the ground to LiDAR measurements is correctly specified for the area of interest for the estimator to be unbiased while the model-assisted estimator is approximately design-unbiased.

Gregoire et al. (2011) and Gobakken et al. (2012) presented two case studies in which a systematic sample of parallel scanning airborne LiDAR flight-lines was collected over Hedmark County, Norway. The size of the county is approximately 27,390 km² and flight-lines were spaced 6 km apart. The LiDAR swath width was 500 m. Thus, approximately 8% of the population was sampled with LiDAR. The LiDAR was flown over the permanent sample of National Forest Inventory (NFI) field plots. The NFI plots are located on a systematic 3 km × 3 km grid. Results seemingly indicated that estimates of biomass per hectare for the entire County and for individual cover classes based on the field survey only, i.e., a so-called direct estimate assuming simple random sampling, was equally precise or even more precise (smaller standard error) than the corresponding estimates based on the model-assisted estimator assuming a two-stage cluster design. This is somewhat surprising since one would expect a sample of LiDAR data with 8% coverage of the entire population to improve the precision considerably.

In a simulated sampling study based on an artificial population of above-ground biomass in Hedmark County, Ene et al. (2012) tested the analytical standard error estimators derived by Gregoire et al. (2011) under systematic sampling and compared the analytical estimates against the empirical estimates obtained by simulations. They also compared the empirical standard error estimates against those obtained for the pure field-based survey. The major findings were that (1) the empirical (“true”) standard error of the LiDAR-assisted cluster design under systematic sampling was only 21% of that obtained with the analytical estimator (i.e., a serious overestimation of the uncertainty with the analytical estimator) and (2) that the empirical standard error of the LiDAR-assisted design was 41% of that obtained from the field sample alone. A relative standard error of 41% translates to a relative variance, also known as relative efficiency, of 5.8, indicating that 5.8 times as many field samples would be required for the pure field-based survey to obtain the same precision as with the LiDAR-assisted survey.

Two other model-assisted studies with scanning LiDAR conducted in two other regions in Norway with complete coverage of LiDAR data for the entire populations in question indicated standard errors of a magnitude of 40–45% and 42% of what was obtained by a pure field survey for above-ground biomass (Næsset et al., 2011) and timber volume (McRoberts et al., 2013), respectively, which correspond well with what Ene et al. (2012) found.

There are however, two important properties of the design in the Hedmark County survey that were taken into consideration in the simulation study by Ene et al. (2012) using an artificial population but which are difficult to account for and cannot be ignored in operational cases like those described by Gregoire et al. (2011) and Gobakken et al. (2012). First, a simple random sample was assumed in both stages (the LiDAR flight-lines and the field sample plots) under the LiDAR-assisted design while the samples were truly systematic in both stages. It is well known that variance estimators assuming random designs are biased under systematic designs. The magnitude of this bias is unknown, although the bias is almost always positive, i.e., the analytical estimates of variance exceed the actual variance. Ene et al. (2012) provided some empirical evidence for this bias for the given artificial population they used. Second, as noted by Gregoire et al. (2011), there are fundamental differences between the two designs (i.e., LiDAR-assisted cluster design and pure field sample) and they rest on different assumptions. One of these differences is that while the sampling variability in simple random
sampling is attributed to variation between the primary sampling units (the individual plots), uncertainty in two-stage sampling is quantified by the variability between sampling units (plots) within clusters as well as between groups (clusters) of plots. These designs are therefore difficult to compare.

As there is an urgent need to come up with scientifically based recommendations for efficient designs to monitor under REDD, it is important to have realistic and comparable values for expected uncertainties and associated inventory costs for alternative designs. A comparison between an estimator assuming simple random sampling and the one developed by Gregoire et al. (2011) could be achieved through simulations. That might address the issue of random versus systematic designs as well as the issue of differences between the estimators as such, as demonstrated by Ene et al. (2012). Nonetheless, a simulation would have to assume a certain model population, which indeed would differ from the true but unknown population of Hedmark County. For studies of uncertainty in sampling, correctly representing spatial correlation is crucial. The latter task is extremely difficult. First, the computational complexity and intensity is very demanding for even relatively small areas. Second, spatial correlation often is not constant and often has directional components. Third, although the task is difficult for relatively homogeneous, non-fragmented forest, it is even more difficult for the fragmented forests like in Hedmark County. Thus, the actual real-world sample data collected in Hedmark offer some opportunities for analysis where the true spatial structure of the population is maintained. Therefore, simulation studies and estimation based on a true sample from a real population are complementary exercises. By assuming more similar designs for the field-based direct estimation and the model-assisted estimation and thus ruling out some of the unequal assumptions that a comparison of simple random sampling estimation and two-stage model-assisted estimation necessarily will suffered from (Andersen et al., 2009; Gobakken et al., 2012; Gregoire et al., 2011), it should be possible to provide estimates of uncertainty that are more comparable and could give an indication of the relative efficiency of these two basic approaches to estimation.

Two alternative strategies for such a comparative analysis exist, namely (1) to adopt designs for the field survey which are more similar to the two-stage cluster design assumed in the model-assisted LiDAR estimation, and (2) to adopt designs for the LiDAR-assisted survey which are more similar to the simple random sampling assumed in the direct field-based estimation. The objective of this study was to follow these two strategies in order to shed further light on the relative magnitude of uncertainties in above-ground biomass estimates (Mg ha\(^{-1}\)) for pure field-based sample surveys and LiDAR-assisted surveys with special reference to Hedmark County. The major findings in this study provide a useful insight of general value that goes beyond previous knowledge and clear recommendations are given regarding the direction of future research to find appropriate design-based and model-assisted estimators for surveys involving LiDAR sampling. Finally, some options for design of LiDAR-assisted sample surveys under REDD are discussed, which capitalize on the flexibility offered when the field survey is designed as an integrated part of the overall survey design as opposed to previous LiDAR-assisted sample surveys in the boreal and temperate zones which have been restricted by the current design of an existing NFI.

2. Material and methods

2.1. Study area

The area of interest (AOI) in this study was Hedmark County (HC). HC is located in southeastern Norway on the Swedish border (Fig. 1) and the total area is approximately 27,390 km\(^2\) with altitude ranging from 119 to 2178 m a.s.l. There is a general trend of rising elevations and thus decreasing productivity from south to north. The dominant tree species are Norway spruce (Picea abies (L.) Karst.) and Scots pine (Pinus sylvestris L.) with extensive tracts of Downy birch (Betula pubescens Ehrh.) close to the mountains.

2.2. Field sample survey

A sub-sample of the permanent field plots of the Norway National Forest Inventory (NFI) was used in the estimation. The NFI plots are located on a 3 km × 3 km grid and each year 20% of the plots are re-measured according to a Latin square design. In the current study, we used approximately 30% (659 plots) of the NFI plots in HC, namely those measured in 2005, 2006, and 2007 and located along the scanning LiDAR flight-lines. We used data from these three years only since that would correspond fairly well to the time of LiDAR data acquisition. The LiDAR flight-lines were flown as parallel strips and located 6 km apart rather than 3 km apart. Had the LiDAR flight-lines been collected every 3 km, then the entire NFI ground sample would have been measured by the LiDAR. However, this expense could not be covered by the project. The NFI plots constitute an un-stratified systematic sample. We divided HC into eight mutually exclusive cover classes, i.e., the four productive cover classes (1) High, (2) Medium, (3) Low productivity forests, and (4) Young forest, and the three nonproductive or non-forest cover classes (5) Nonproductive forest, (6) Mountain areas, (7) Open water, and (8) Developed areas. Cover class Developed areas was excluded from this study because the field sample was not selected according to probability-based principles. The classes were formed on the basis of existing official land use maps combined with classification of Landsat TM data. Detailed definitions of the seven non-overlapping cover classes and the composite cover class map that was produced can be found in Gobakken et al. (2012).

The NFI plots are circular with size 250 m\(^2\). On each sample plot, all trees with diameter at breast height ≥ 5 cm were callipered and tree heights were measured on an average of 10 sample trees per plot. Total above-ground dry forest biomass (AGB, Mg ha\(^{-1}\)) was computed as the sum of the individual biomass components, i.e., stump, stem, bark, dead and living branches, and foliage of all individual trees, using species-specific allometric models (Marklund, 1988) with diameter at breast height and tree height as independent variables following the procedures described in Gobakken et al. (2012). The plot positions were determined by differential post-processing of dual-frequency observations of the American Global Positioning System (GPS) and the Russian Global Navigation Satellite System, with estimated accuracy of the plot coordinates ranging from 0 to 2 m, with an average of 0.05 m (Gobakken et al., 2012; Nasset, 2001).

An overview of the data is presented in Table 1. There seems to be a geographical trend of increasing AGB values in the NFI sample from north to south (Fig. 2).

2.3. Additional data from forest management inventories in Hedmark County

The short-range (< 3 km) spatial correlation of AGB was a concern in some of the analyses. We therefore used data from six LiDAR-assisted forest management inventories in Hedmark to help with a simple assessment of a likely range of the spatial correlation in AGB. The forest in HC is generally actively managed according to standard silvicultural practices typically seen in boreal forests. Stands are typically harvested by clear-felling, and planting, tending, and thinning are treatments frequently seen in many of the stands. Thus, the forest landscape has a pronounced stand structure with homogenous and single-aged stands, often consisting of mono-cultures. In general, a strong spatial correlation in AGB can therefore be found within stands while the stand boundaries often indicate abrupt changes in the distributional patterns of AGB. The typical stand size in HC can therefore give an indication of the geographical scale at which AGB can be expected to display a strong correlation. Examination of stand sizes in the six inventories (Fig. 1) covering a total area of 1102.4 km\(^2\), showed an average stand size ranging...
were conducted. The six areas are marked with numbers (see Table 2 for further details). A detailed close-up from Hedmark County (right) displays Gray-shaded background indicates above-ground biomass density (dark is high biomass; light is low biomass). Black areas indicate where the forest management inventories were conducted. The six areas are marked with numbers (see Table 2 for further details). A detailed close-up from Hedmark County (right) displays five of the LiDAR flight-lines with their corresponding swaths and National Forest Inventory field plots (dots).

from 0.87 ha in Area 2 (Sør-Odal Municipality) to 2.07 ha in Area 5 (Åmot Municipality (east)) (Table 2). By assuming square-shaped stands, the maximum range for an average stand would be around 93–144 m. A crude estimate of range of spatial correlation of AGB taking into account that stands in reality seldom are geometric units as compact as a square could therefore be, say, 200 m.

2.4. Airborne scanning LiDAR data

The LiDAR data were acquired in the period between 22 July and 16 September 2006. Fifty-three parallel flight-lines were flown with an inter-line distance of 6 km (Fig. 1). Fixed-winged aircrafts were used to carry the Optech ALTM 3100 laser scanning systems (Optech, Canada). Average flying altitude was approximately 800 m a.g.l. at a flight speed of 75 ms⁻¹. The pulse repetition frequency was 100 kHz and the scan frequency was 55 Hz. A maximum half scan-angle of 17° resulted in a swath width of approximately 500 m. The average footprint diameter was 21 cm and the average pulse density was 2.8 m⁻².

The data were processed by the contractor (Blom Geomatics, Norway). Ground echoes were found and classified using the progressive classification algorithm (Axelsson, 2000) of the TerraScan software (Anonymous, 2005). Heights above the created TIN surface were calculated for all echoes by subtracting the respective TIN heights from the height values of all echoes recorded. The ALTM 3100 sensor is capable of recording up to four echoes per pulse. Echoes classified as “single” and “first of many” were pooled into one dataset denoted as “first” echoes. Similarly, echoes classified as “single” and “last of many” were pooled into another dataset denoted as “last” echoes. The first and last datasets were stored for subsequent estimation.

The 500 m wide LiDAR strips were divided into regular cells with size 250 m² (Fig. 3) and each cell was assigned to its corresponding cover class. For each cell, separate canopy height distributions were extracted for those LiDAR echoes of the first and last echo datasets, respectively, that exceeded 2 m above the ground surface. The LiDAR data were also laid atop the NFI field plots and the corresponding canopy height distributions were extracted for the field plots as well, i.e., for LiDAR echoes inside the plot circumference. For every grid cell and every field plot we derived height-related metrics from the height distributions, such as height deciles. Further, we also derived variables related to canopy density, such as the relative cumulative densities at various vertical height levels, see Gobakken et al. (2012) for details. These LiDAR-derived metrics were used in the subsequent estimation.

2.5. Estimation and inference

In order to more fairly compare uncertainties of ground-based and LiDAR-assisted above-ground biomass estimates and thus rule out effects of unequal assumptions, a design for model-assisted estimation with LiDAR as auxiliary information was sought that mimicked, as best we could, the design of the field survey. Since simple random sampling was assumed for the field survey, a two-phase (double sampling) approach was employed for the model-assisted LiDAR estimation. This approach effectively forces us to handle the LiDAR-assisted survey in a fashion similar to the field survey.

Conversely, we can “force” the field survey to look like a LiDAR-assisted design. Gregoire et al. (2011) assumed a two-stage (cluster)

Table 1

<table>
<thead>
<tr>
<th>Cover class</th>
<th>Area (%)</th>
<th>No. of NFI plots</th>
<th>AGB (Mg ha⁻¹)</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productive forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>48</td>
<td>123.9</td>
<td>0.0</td>
<td>311.5</td>
</tr>
<tr>
<td>Medium</td>
<td>13</td>
<td>105</td>
<td>96.7</td>
<td>4.9</td>
<td>290.9</td>
</tr>
<tr>
<td>Low</td>
<td>16</td>
<td>141</td>
<td>49.3</td>
<td>0.0</td>
<td>177.5</td>
</tr>
<tr>
<td>Young</td>
<td>17</td>
<td>151</td>
<td>33.2</td>
<td>0.0</td>
<td>207.0</td>
</tr>
<tr>
<td>Nonproductive forest and nonforest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonproductive forest</td>
<td>11</td>
<td>83</td>
<td>28.0</td>
<td>0.0</td>
<td>151.2</td>
</tr>
<tr>
<td>Mountain areas</td>
<td>28</td>
<td>95</td>
<td>12.4</td>
<td>0.0</td>
<td>128.8</td>
</tr>
<tr>
<td>Open water</td>
<td>5</td>
<td>36</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Triangular Irregular Network (TIN) densification algorithm (Axelsson, 2000) of the TerraScan software (Anonymous, 2005). Heights above the created TIN surface were calculated for all echoes by subtracting the respective TIN heights from the height values of all echoes recorded. The ALTM 3100 sensor is capable of recording up to four echoes per pulse. Echoes classified as “single” and “first of many” were pooled into one dataset denoted as “first” echoes. Similarly, echoes classified as “single” and “last of many” were pooled into another dataset denoted as “last” echoes. The first and last datasets were stored for subsequent estimation.

The 500 m wide LiDAR strips were divided into regular cells with size 250 m² (Fig. 3) and each cell was assigned to its corresponding cover class. For each cell, separate canopy height distributions were extracted for those LiDAR echoes of the first and last echo datasets, respectively, that exceeded 2 m above the ground surface. The LiDAR data were also laid atop the NFI field plots and the corresponding canopy height distributions were extracted for the field plots as well, i.e., for LiDAR echoes inside the plot circumference. For every grid cell and every field plot we derived height-related metrics from the height distributions, such as height deciles. Further, we also derived variables related to canopy density, such as the relative cumulative densities at various vertical height levels, see Gobakken et al. (2012) for details. These LiDAR-derived metrics were used in the subsequent estimation.

2.5. Estimation and inference

In order to more fairly compare uncertainties of ground-based and LiDAR-assisted above-ground biomass estimates and thus rule out effects of unequal assumptions, a design for model-assisted estimation with LiDAR as auxiliary information was sought that mimicked, as best we could, the design of the field survey. Since simple random sampling was assumed for the field survey, a two-phase (double sampling) approach was employed for the model-assisted LiDAR estimation. This approach effectively forces us to handle the LiDAR-assisted survey in a fashion similar to the field survey.

Conversely, we can “force” the field survey to look like a LiDAR-assisted design. Gregoire et al. (2011) assumed a two-stage (cluster)
design for the model-assisted estimation with LiDAR. A more comparable design for direct estimation based on the field survey than simple random sampling would be two-stage equal probability cluster sampling.

Thus, we have identified four different designs:

1. Simple random sampling — direct estimation based on the field survey only.
2. Two-stage (cluster) sampling — direct estimation based on the field survey only.
3. Two-phase (double) sampling — model-assisted estimation supported by LiDAR data as auxiliary information.
4. Two-stage (cluster) sampling — model-assisted estimation supported by LiDAR data as auxiliary information as per Gregoire et al. (2011).

A common property of the four designs mentioned above and the corresponding estimators that we employed for each of them (see below) is that they are so-called design-based estimators. All these estimators rely on probability sampling, i.e., that there are known and positive probabilities for selection for each element in the defined population and that there exist one and only one positive value for each population element. As illustrated in Fig. 3, a basic property of two-phase sampling is that a more intense, fine-scale (small sampling units) sample is selected in the first phase with a less intense subsample selected in the second phase, as opposed to two-stage sampling where a less intense, coarse-scale (large sampling units; clusters) sample is selected initially, and a more intense, finer-scale sample is selected subsequently within the first-stage sampling units. It is a common property of all designs though that the selection of sampling units is random. That even applies to the two-phase and two-stage designs where random selection of sampling units (individual population elements (plots) in two-phase and individual clusters in two-stage) is assumed in the first phase and stage of sampling with subsequent random selection of individual population elements (plots) within the first-phase elements and first-stage clusters, respectively.

In the following, we want to provide appropriate estimators for AGB and corresponding error estimators for the four identified designs.

2.5.1. Simple random sampling — direct estimation (SRS)

We find it convenient first to detail the estimation for the cover classes. Let \( U \) be the entire population of elements (grid cells with size 250 m\(^2\)) in HC where \( U = \{1,...,N\} \). The population was divided into non-overlapping cover classes, \( U_c \), with sizes \( N_c \), where \( c = 1,...,H \).

Now, let \( b_k \) be the total above-ground biomass per hectare of the \( k \)-th element in the population. We want to define the parameter mean biomass per hectare (\( \bar{B} \)) within a particular cover class (\( c \)) for which we later wish to find an appropriate estimator:

\[
\bar{B}_c = \frac{\sum b_k}{N_c}.
\]

(1)

First, we want to estimate the above-ground biomass per hectare from the field sample alone, i.e., using a so-called direct estimator. Let \( s_c \) denote a sample of size \( n_c \) that after random selection [i.e., simple random sampling (SRS)] falls in cover class \( c \). Then \( s_c \) is the subsample of the plots in \( U_c \). The mean above-ground biomass per hectare for a particular cover class was estimated according to (Gregoire & Valentine, 2008):

\[
\hat{B}_{acs} = \frac{\sum b_k}{n_c}.
\]

(2)
This estimator will be slightly biased since it is a ratio estimator due to the random sample size, \( n_c \).

An estimator for the variance of \( \hat{B}_{\text{SRS}} \) is

\[
\hat{V}(\hat{B}_{\text{SRS}}) = \frac{\sum_{k \in \mathcal{S}} (b_k - \hat{B}_{\text{SRS}})^2}{\frac{n_c(n_c - 1)}{n_c}}.
\]  

(3)

In this estimator we have ignored corrections for finite population because the sampling fractions are always very small and a correction would have a negligible influence on the variance estimates.

In situations where a single estimate was sought across several cover classes, e.g., for cover classes \( U_1 \) and \( U_2 \), we applied the estimators in Eqs. (2) and (3) by collapsing the two classes and let them be represented by the same subscript \( c \). For estimation of overall above-ground biomass per hectare in HC, we ignored the cover classes entirely. Thus, based on the overall sample \( s \) of size \( n \) biomass was estimated according to

\[
\hat{B}_{\text{SRS}} = \frac{\sum_{k \in \mathcal{S}} b_k}{n}
\]  

(4)

with the variance estimator

\[
\hat{V}(\hat{B}_{\text{SRS}}) = \frac{\sum_{k \in \mathcal{S}} (b_k - \hat{B}_{\text{SRS}})^2}{\frac{n(n-1)}{n}}.
\]  

(5)

### 2.5.2. Two-stage (cluster) sampling — direct estimation (2ST)

In the following we will assume that the population \( U \) is partitioned into \( M \) clusters denoted \( U_1, \ldots, U_c, \ldots, U_{m_c} \). The clusters are non-overlapping and each of them is assumed to have the same shape and size as the swath produce by the scanning LiDAR along an individual flight-line. The number of population elements (grid cells with size 250 m\(^2\)) within cluster \( U_i \) is denoted \( N_i \). We will now assume that \( m \) among the \( M \) clusters will be selected by simple random sampling. This sample of clusters \( \{s_1\} \) constitutes our first-stage sample. For every first-stage cluster \( U_i \) we will then select a second-stage sample \( s_1 \) of size \( n_i \) among the \( N_i \) population elements in \( U_i \) according to simple random sampling.

An appropriate estimator for mean above-ground biomass per hectare for HC for the two-stage design (2ST) is (Gregoire & Valentine, 2008, p. 397)

\[
\hat{B}_{\text{2ST}} = \frac{M}{N_m} \sum_{i \in s_1} \frac{N_i}{N_m} \sum_{k \in s_2} b_k.
\]  

(6)

An estimator for the variance of \( \hat{B}_{\text{2ST}} \) is

\[
\hat{V}(\hat{B}_{\text{2ST}}) = \frac{N_m}{M^2} \sum_{i \in s_1} \left( \frac{1}{m} - \frac{1}{M} \right) s_i^2 + \frac{M}{m} \sum_{i \in s_1} N_i \left( \frac{1}{n_i} - \frac{1}{N_i} \right) \hat{S}_i^2
\]  

(7)

where

\[
S_i^2 = \frac{1}{m-1} \sum_{k \in s_2} \left( b_k - \frac{1}{n_i} \sum_{k \in s_2} b_k \right)^2
\]  

(8)

and

\[
\hat{S}_i^2 = \frac{1}{n_i-1} \sum_{k \in s_2} \left( b_k - \frac{1}{n_i} \sum_{k \in s_2} b_k \right)^2
\]  

(9)

Eq. (8) quantifies the variation between individual cluster totals of above-ground biomass and the mean total biomass over all clusters. Eq. (9) quantifies the variation between above-ground biomass per hectare for individual plots within a given cluster \( i \) and mean biomass over all plots for that particular cluster.

Further, we wish to estimate above-ground biomass per hectare for cover class \( c \). The number of population elements within cluster \( U_i \) that falls in class \( c \) is denoted \( N_{ic} \). The second-stage sample that falls in class \( c \) is denoted \( s_{ic} \) and is of size \( n_{ic} \). An appropriate estimator for cover class \( c \) is obtained by introducing the subscript \( c \) in Eq. (6):

\[
\hat{B}_{\text{2STc}} = \frac{M}{N_m} \sum_{i \in s_1} \frac{N_{ic}}{N_m} \sum_{k \in s_{ic}} b_k
\]  

(10)

while, similarly, an estimator for the variance of \( \hat{B}_{\text{2STc}} \) is

\[
\hat{V}(\hat{B}_{\text{2STc}}) = \frac{N_m}{M^2} \sum_{i \in s_1} \left( \frac{1}{m} - \frac{1}{M} \right) s_i^2 + \frac{M}{m} \sum_{i \in s_1} N_i \left( \frac{1}{n_i} - \frac{1}{N_i} \right) \hat{S}_{ic}^2
\]  

(11)
where

\[ S_2^2 = \frac{1}{m-1} \sum_{u \in U_{SC}} \left( \frac{N_2}{n_2} \sum_{k=1}^{n_2} b_k - \frac{N_2}{M} \beta_{2SC} \right)^2 \]

(12)

and

\[ S_1^2 = \frac{1}{n_1-1} \sum_{k=1}^{n_1} \left( b_k - \frac{1}{n_1} \sum_{k=1}^{n_1} b_k \right)^2. \]

(13)

In situations where a single estimate was sought across several cover classes, the classes in question were collapsed and represented by the same subscript \( c \) in Eqs. (10)–(13).

2.5.3. Two-phase (double) sampling — model-assisted estimation (2PHMA)

We will now assume for analytical purposes that a large sample \( s_1 \) of size \( n_1 \) is selected by simple random sampling among the \( N \) population elements (grid cells with size 250 m\(^2\)) in \( U \). The sample \( s_1 \) constitutes our first-phase sample. For all elements in \( s_1 \) we have auxiliary data from the LiDAR at hand. The cluster structure assumed for the LiDAR data in the two-stage design is therefore irrelevant in this setup. Further, we wish to select a second and smaller sample by random selection among the \( n_1 \) elements in the larger first-phase sample \( s_1 \). This smaller second-phase sample is denoted \( s_2 \) and is of size \( n_2 \). The second-phase sample \( s_2 \) is our field sample. Thus, for \( s_2 \) we have estimated above-ground biomass from field observations as well as auxiliary data from the LiDAR. Note that in the comparison of estimates we will assume that \( s_2 \) is identical to the sample \( s \) under simple random sampling (see above).

Based on the second-phase sample \( s_2 \), we will now fit a regression model that relates the above-ground biomass per hectare to the LiDAR-derived metrics. Alternatively, one may also choose to adopt a biomass-LiDAR model fitted to another dataset or taken from another geographical area. The regression model we chose to adopt will be used to predict above-ground biomass per hectare for the elements in \( s_1 \).

An appropriate model-assisted regression estimator for mean above-ground biomass per hectare for HC for the two-phase design (2PHMA) is (Mandallaz, 2008, p. 80; Särndal et al., 1992, p. 364)

\[ \hat{B}_{2PHMA} = \frac{\sum_{k=1}^{n_2} b_k}{n_1} + \frac{\sum_{k=1}^{n_2} e_k}{n_2} \]

(14)

where \( b_k \) is biomass per hectare predicted according to the regression model for the \( k \)th element in the population and \( e_k = b_k - b_k \). This estimator is approximately design-unbiased irrespective of the model choice when \( n_2 \) is not too small (Särndal et al., 1992). Firth and Bennett (1998) established that asymptotic design-unbiasedness holds for nonlinear model-assisted regression, as well.

An estimator for the variance of \( \hat{B}_{2PHMA} \) is (Mandallaz, 2008, p. 81)

\[ \hat{V}(\hat{B}_{2PHMA}) = \left(1 - \frac{n_2}{n_1}\right) \frac{\sum_{k=1}^{n_2} e_k^2}{n_2(n_2-1)} + \frac{\sum_{k=1}^{n_2} (b_k - \frac{1}{n_2} \sum_{k=1}^{n_2} b_k)^2}{n_1(n_1-1)}. \]

(15)

Also when assuming a two-phase design we wish to estimate above-ground biomass per hectare for cover class \( c \). Let \( s_{1c} \) denote a sample of size \( n_{1c} \) that after the random selection in the first phase falls in class \( c \). Further, let \( s_{2c} \) denote a sample of size \( n_{2c} \) that after the random selection in the second phase falls in class \( c \). Note that in the comparison of estimates we will assume that \( s_{2c} \) is identical to the sample \( s \) under simple random sampling (see above). An appropriate estimator for cover class \( c \) is obtained by introducing the subscript \( c \) in Eq. (14), which then becomes

\[ \hat{B}_{2PHMAC} = \frac{\sum_{k=1}^{n_{2c}} b_k}{n_{1c}} + \frac{\sum_{k=1}^{n_{2c}} e_k}{n_{2c}} \]

(16)

while, similarly, an estimator for the variance of \( \hat{B}_{2PHMAC} \) is

\[ \hat{V}(\hat{B}_{2PHMAC}) = \left(1 - \frac{n_{2c}}{n_{1c}}\right) \frac{\sum_{k=1}^{n_{2c}} e_k^2}{n_{2c}(n_{2c}-1)} + \frac{\sum_{k=1}^{n_{2c}} (b_k - \frac{1}{n_{2c}} \sum_{k=1}^{n_{2c}} b_k)^2}{n_{1c}(n_{1c}-1)}. \]

(17)

As with the estimators for the other designs, estimates across several cover classes will be provided by collapsing the classes in question and assigning the same subscript \( c \) in Eqs. (16) and (17) to the respective classes.

2.5.4. Two-stage (cluster) sampling — model-assisted estimation (2STMA)

In the two-stage (cluster) design for model-assisted estimation, it is assumed that the population \( U \) is partitioned into \( M \) clusters (primary sampling units) in exactly the same way as for two-stage design for direct estimation. The clusters are non-overlapping and each of them is assumed to have the same shape and size as the swath produced by the scanning LiDAR along an individual flight-line. Further, it is assumed that each of the \( M \) clusters will be selected by simple random sampling, and the sample of \( M \) clusters is our first-stage sample. In exactly the same way as with the two-stage design treated in the previous section, a second-stage sample will be selected among the population elements within each of the first-stage clusters according to simple random sampling. However, as opposed to the direct estimation, the model-assisted estimation will take advantage of the auxiliary LiDAR data available for all population elements within the first-stage clusters.

This particular design is treated by Särndal et al. (1992, p. 323). Gregoire et al. (2011) elaborated further on the estimators provided by Särndal et al. (1992). We will therefore refer the reader to the estimators presented in Gregoire et al. (2011), specifically their Eqs. (5), (6), (7), (28) and (29). Gobakken et al. (2012) applied these estimators to the particular survey conducted in HC. The estimation will not be repeated in the current work. In this article, we will therefore refer to the estimates provided by Gobakken et al. (2012).

2.6. Analysis

2.6.1. Estimation of LiDAR biomass models

Regression models that relate the LiDAR variables to above-ground biomass per hectare are required for the model-assisted estimation. Since Gobakken et al. (2012) already had fitted such models for the respective cover classes based on the current sample of NFI plots in HC, we applied these previously fitted models rather than estimating the same models once more. Further details reading the regression analysis can be found in Gobakken et al. (2012). However, these models are presented also in the current article (Table 3). For the cover class Open water we did not have a separate model at hand because all NFI plots in this class had AGB = 0.

The models had been fitted as linear regressions in the logarithmic transformations of the variables, i.e.,

\[ \ln Y = \ln f_0 + X\beta + e \]

(18)

where \( Y \) = the AGB field values, \( \beta \) is a vector of regression coefficients, \( X \) is a matrix of the explanatory variables at log scale, and \( e \) is a normally distributed error term. In the back-transformation of the fitted models to arithmetic scale, the parameter estimate of \( f_0 \) was corrected for bias by a ratio correction (Snowdon, 1991).
2.6.2. Estimation of mean biomass and variance of the mean biomass estimates

Assessment of uncertainties associated with comparable designs with and without utilizing the LiDAR data as auxiliary information was the primary objective of this study. The main emphasize in the following is therefore on the estimated standard errors (SE), i.e., the square roots of the variances.

First, we estimated above-ground biomass per hectare for every cover class and groups of cover classes (e.g. productive forest) (Eq. 2) and for entire HC (Eq. 4) based on the field sample only. We also estimated the corresponding standard errors (square roots of the variances in Eqs. 3 and 5, respectively). Because a systematic design was adopted for the field survey rather than a random design, an overestimation of the variance is a likely consequence of ignoring the systematic design (e.g. Särndal et al., 1992).

Second, above-ground biomass per hectare was estimated for the same cover classes, groups of cover classes, and entire HC assuming the two-stage design for direct estimation (Eqs. 10 and 6, respectively). The first-stage sampling units were assumed to be E–W oriented clusters spaced 6 km apart and with a width of 500 m. Corresponding standard errors were estimated according to Eqs. (11) and (7), respectively. Because there is tendency of a geographical trend in AGB with increasing values as one goes from north to south (Fig. 2), orientation of the clusters might influence on the magnitude of the sampling uncertainty. Therefore, we repeated the estimation assuming a two-stage design with N–S oriented clusters. In this case we assumed parallel clusters with width equal to 500 m spaced 3 km apart.

Third, we estimated above-ground biomass per hectare for individual cover classes, groups of cover classes, and entire HC assuming a two-phase design (Eqs. 16 and 14, respectively). The design assumes independent random samples in both phases. However, we truly selected our samples systematically in both phases, i.e., with first-phase sample units at fixed intervals within the LiDAR swaths (the swaths were parallel and spaced 6 km apart) and the second-phase sample units at fixed intervals of 3 km (i.e., the NFI plots) along the swaths. It was essential to select a minimum geographical distance between spatially adjacent first-phase sample units that would avoid serious spatial correlation in the AGB values. Based on the assessment of the average stand sizes in six selected forest management inventories in HC (see above), we chose a first-phase sampling interval of 200 m. The principles of the systematic selection of first- and second-phase sample units along a LiDAR swath is illustrated in Fig. 3. In the model-assisted estimation, above-ground biomass per hectare was predicted for the first-phase sample units, i.e., the 250 m² grid cells, using the cover class-specific regression models (Table 3). For the cover class Open water, however, we used the model fitted for Medium productivity forest since we did not have a separate model for Open water at hand, which is in correspondence with Gobakken et al. (2012). Field plots as well as population elements with LiDAR observations indicating positive biomass values sometimes occur in Open water because of inaccuracies in the map data used to define the cover classes. Even when we estimated above-ground biomass for entire HC and ignored cover classes (Eq. 14) the specific models were applied to the first-phase sample units in accordance with their respective cover class assignment. Standard errors of the above-ground biomass estimates per hectare were estimated according to Eqs. (17) and (15), respectively.

Finally, the estimates from implementation of the two-stage design with model-assisted estimation were taken from Gobakken et al. (2012). In this estimation, the E–W oriented and parallel LiDAR swaths with a width of 500 m and spaced 6 km apart were considered the first-stage sample units. The second-stage sample was constituted by the NFI plots spaced 3 km apart within the first-stage units. All population elements (grid cells with size 250 m²) within the LiDAR swaths were used in the model-assisted estimation. The two-stage and two-phase model-assisted estimation followed the same principles as far as usage of the cover class specific regression models is concerned. The four different designs are illustrated graphically in Fig. 3.

3. Results and discussion

3.1. SRS direct estimation versus two-phase model-assisted estimation (2PHMA)

Most of the cover class-specific means above-ground biomass estimates obtained assuming the SRS and 2PHMA designs were of a similar magnitude. The largest absolute difference was found for High productive forest with an estimated mean biomass of 123.9 and 133.3 Mg ha⁻¹ for SRS and 2PHMA, respectively (Table 4). For Open water none of the NFI sample plots carried any biomass, thus the SRS estimate was zero, while the model-assisted mean estimate was 4.3 Mg ha⁻¹. As noted above, field plots as well as population elements with LiDAR observations indicating positive biomass values sometimes occur in Open water because of inaccuracies in the map data used to define the cover classes. The overall mean above-ground biomass estimates across all cover classes were 39.6 and 41.4 Mg ha⁻¹ for SRS and 2PHMA, respectively.

The estimated standard errors deviated considerably between the two designs within individual cover classes. For all classes the standard error estimates were 1.7 and 0.6 Mg ha⁻¹ for SRS and 2PHMA (Table 4), respectively, which translates to a relative efficiency of 80. This is somewhat higher than the efficiency of 5.8 reported by Ene et al. (2012) for two-stage model-assisted against SRS and 5–6 reported by Næsset et al. (2011) and McRoberts et al. (2013). It should be noted though that although the estimation across all cover classes was unstratified, we allowed class-specific models to be used in the individual classes (see details above). By inspecting the relative efficiency within each individual cover class, it can be seen that the overall efficiency was higher than those within the individual classes (5.0–7.9). Nevertheless, 2PHMA consistently produced lower standard error estimates than SRS for every cover class, regardless of the specificities of the individual classes. By allowing for only one joint model across all cover classes or eventually a stratified estimation for the SRS design as well as for 2PHMA, it is reasonable to expect a somewhat lower efficiency than 80. On the other hand, the two-phase model-assisted design assumed only one first-phase plot every 200 m along the LiDAR flight-lines. Thus, 99.75% (399/400) of the continuously collected LiDAR data were discarded under the 2PHMA design. Although there is a strong spatial correlation of AGB

<table>
<thead>
<tr>
<th>Cover class</th>
<th>Predictor variables*</th>
<th>n²</th>
<th>R²</th>
<th>RMSE</th>
<th>RMSEannon</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productive forest</td>
<td>lnhmean, ln1L, lnAlt</td>
<td>46</td>
<td>0.94</td>
<td>0.17</td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>lnhmean, ln1L, lnAlt</td>
<td>105</td>
<td>0.95</td>
<td>0.19</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>lnhmean, ln1L, lnAlt</td>
<td>138</td>
<td>0.89</td>
<td>0.36</td>
<td>26.8</td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>ln1L, lnhmean, ln1L</td>
<td>151</td>
<td>0.86</td>
<td>0.50</td>
<td>44.9</td>
<td></td>
</tr>
</tbody>
</table>

| Nonproductive forest | ln1L, ln1L, lnAlt | 107 | 0.80 | 0.60 | 45.3 |
| Mountain areas | ln1L, ln1L | 85 | 0.87 | 0.46 | 37.0 |

<table>
<thead>
<tr>
<th>Notes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a H100 and h50 are percentiles of the first echo LiDAR canopy heights for 50% and 90% (m); hmean = arithmetic mean of first or last echo LiDAR canopy heights, respectively (m);</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d100 = canopy density corresponding to the proportion of first echo LiDAR echoes above fraction #1 to total number of first echoes (see Gobakken et al. (2012)); and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d100 = canopy densities corresponding to the proportions of last echo LiDAR echoes above fraction #1 to total number of last echoes; Alt = altitude (m a.s.l.).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b Numbers of plots differ from those in Table 1 because plots with AGB = 0 were discarded from the regression model fitting.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c RMSE of the back-transformed predicted AGB versus field measured AGB. RMSE is given in percent of mean AGB.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
over short distances and thus a limited marginal utility of continuously
collected LiDAR data, there is no doubt that LiDAR data collected
continuously along a flight-line and with a swath width of 500 m as
in the current study, will improve precision over the employed
2PHEMA design. Thus, the gain in precision obtained in this study by
borrowing support from the auxiliary LiDAR data seems to be fairly
well in line with results of those previous studies that have analyzed
comparable LiDAR-assisted designs against pure field surveys.

In a recent study from New Zealand conducted in conifer forests
where airborne scanning LiDAR was used in a classical double-sampling
setup with regression estimation, a gain in precision of 6% for carbon
stock was reported (Stephens et al., 2012) when the first-phase sample
size was only 1.19 times larger than in the second-phase sample. In our
study, this ratio was 15. They stated that the gain in precision over simple
random sampling could potentially be up to about 50% by increasing the
first-phase sample size. However, since a scanning LiDAR instrument
typically collects data continuously along a flight-line, exploiting the full
potential of the LiDAR data by employing a design that takes into account all
the auxiliary data, like for example a two-stage cluster design, would
most likely give a substantial improvement of the precision over the po-
tential indicated in the study from New Zealand.

3.2. Two-stage direct estimation (2ST) versus two-stage model-assisted
estimation (2STMA)

The mean above-ground biomass estimates for individual cover
classes as well as for all classes obtained with the 2STMA estimator by
Gobakken et al. (2012) corresponded quite well with the two-stage di-
rect estimates assuming N–S orientation of the first-stage samples. The
overall model-assisted estimate was 38.1 Mg ha\(^{-1}\) whereas the corre-
spending direct estimate under N–S orientation was 39.0 Mg ha\(^{-1}\)
(Table 4). When assuming E–W orientation, however, which also was
the orientation of the LiDAR flight-lines (Fig. 1), the 2ST and 2STMA es-
estimates deviated considerably. For 2ST, the overall mean above-ground
biomass estimate was 46.1 Mg ha\(^{-1}\). The difference between 2ST with
E–W orientation and 2STMA as well as between the two individual 2ST
estimates can be attributed to a large difference for the Mountain cover
class which accounts for 28% of the total area of HC (Table 1). For this
particular cover class the estimates were 5.1 (SE = 0.9), 8.2 (SE = 1.5),
and 18.5 (SE = 5.2) Mg ha\(^{-1}\) for 2STMA, 2ST N–S, and 2ST E–W, respec-
tively (Table 4). Note the large uncertainty (SE = 5.2 Mg ha\(^{-1}\)) in the
estimate that deviated most from the others, namely 2ST E–W.

The standard error of the overall model-assisted mean estimate and the corresponding standard errors for the direct estimates with
E–W and N–S orientation were 1.9, 3.0, and 3.9 Mg ha\(^{-1}\), respectively,
indicating relative efficiencies of 2.5 and 4.2, respectively. These un-
certainty estimates indicate a somewhat smaller gain in precision by
supporting the survey with auxiliary LiDAR data than what has been in-
dicated by the previous studies mentioned above. However, despite the
seemingly smaller gain in precision, all uncertainty estimates consistently
indicate a potential of substantial improvement of precision by extensive
use of LiDAR data. This is in contrast to those previous studies that have
compared analytical uncertainty estimates of pure field surveys assuming
simple random sampling against complex two-stage designs employing
analytical model-assisted estimators (e.g. Andersen et al., 2009; Gobakken et al., 2012; Gregoire et al., 2011; Nelson et al., 2012). It should
be noted that negative variances were experienced for some of the indi-
vidual cover classes when employing the model-assisted variance esti-
mator (Table 4). These negative variances are due to the fact that
within-flight-line variation overwhelms the other variance components
[see Eq. 28 in Gregoire et al. (2011)]. The negative variances are associ-
ated with samples of sizes smaller than what is generally recommended as
a minimum size [n ≥ 5 (Thompson, 2002, p.159); n ≥ 10 (Sundal et al., 1992, p. 407)].

The results also showed a somewhat larger estimated uncertainty
when we assumed N–S orientation of the first-stage samples (SE = 3.9 Mg ha\(^{-1}\)) as compared to E–W orientation (SE = 3.0 Mg ha\(^{-1}\)). It was anticipated that the N–S trend of increasing above-ground biomass
with decreasing latitude (Fig. 2) would be better captured within the
first-stage samples when they were N–S oriented and that an N–S ori-
entation of the samples therefore would perform better. Cluster sam-
ping is generally favorable when the variation is captured within the
clusters leaving little variation between the clusters. However, there is
indeed an N–S trend in biomass in HC, but the variability in biomass be-
tween geographically adjacent plots is still huge, as illustrated by the in-
tervals (± 1 stdev) around the mean above-ground biomass values
over the N–S and E–W lines of NFI plots (Fig. 2).

Furthermore, varying sizes of the first-stage sample units, i.e., varying
lengths of the flight-lines, will according to the analytical variance
estimator have a large impact on the sampling variability. This is due to
the variance component of the first-stage sampling that quantifies the
variability between first-stage sample totals around the common
mean over all first-stage sample totals, see e.g. Eq. (8). The variance es-
timator for the two-stage model-assisted mean above-ground biomass
estimator has an identical construction to account for first-stage

<table>
<thead>
<tr>
<th>Cover class</th>
<th>Simple random sampling</th>
<th>Two-stage</th>
<th>North–south</th>
<th>LiDAR-assisted estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct estimation</td>
<td>Two-stage</td>
<td></td>
<td>Two-phase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>East–west</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productive forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>123.9</td>
<td>98.5</td>
<td>17.6</td>
<td>107.2</td>
</tr>
<tr>
<td>Medium</td>
<td>96.7</td>
<td>90.6</td>
<td>11.8</td>
<td>86.2</td>
</tr>
<tr>
<td>Low</td>
<td>49.3</td>
<td>49.0</td>
<td>5.7</td>
<td>43.8</td>
</tr>
<tr>
<td>Young</td>
<td>33.2</td>
<td>33.0</td>
<td>4.7</td>
<td>31.9</td>
</tr>
<tr>
<td>All productive forests</td>
<td>63.1</td>
<td>64.4</td>
<td>5.6</td>
<td>60.8</td>
</tr>
<tr>
<td>Nonproductive forest and nonforest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonproductive forest</td>
<td>24.4</td>
<td>19.6</td>
<td>3.5</td>
<td>19.5</td>
</tr>
<tr>
<td>Mountain areas</td>
<td>7.4</td>
<td>11.1</td>
<td>18.5</td>
<td>8.2</td>
</tr>
<tr>
<td>Open water(a)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>All n.prod. and nonforest</td>
<td>11.3</td>
<td>15.3</td>
<td>2.7</td>
<td>11.5</td>
</tr>
<tr>
<td>All classes</td>
<td>39.6</td>
<td>46.1</td>
<td>3.0</td>
<td>39.0</td>
</tr>
</tbody>
</table>

\(a\) For the Open water cover class SE = 0 because above-ground biomass was absent on all NFI field plots.

\(b\) Results according to Gobakken et al. (2012). NA denotes the occurrence of a negative estimate of SE.
Those presented by Gregoire et al. (2011) severely overestimate the scanning LiDAR. Nelson et al. (2008) in a simulation study using pro
successive differences estimators and reported that they approximated
rather than simple random sampling could not be assessed. Neverthe-
study. However, in the current study the effect of a systematic design
true variance when the design is truly systematic as in the current
cause of deviation between the analytical variance estimate and the
sampling for the employed 2STMA variance estimator is another major
true variance for uneven cluster sizes and when there is a clear trend
fl
demonstrated by Ene et al. (2012), analytical variance estimators like
obtained by Ene et al. (2012) have indicated that such a design can pro-
more precise estimates than pure field sampling. Ene et al. (in review)
showed by simulations that two-stage sampling with long
flight-lines with uneven lengths was cost-efficient as well. However, as
 demonstrated by Ene et al. (2012), analytical variance estimators like
those presented by Gregoire et al. (2011) severely overestimate the true variance for uneven cluster sizes and when there is a clear trend
in the population. The commonly adopted assumption of simple random
sampling for the employed 2STMA variance estimator is another major
cause of deviation between the analytical variance estimate and the true variance when the design is truly systematic as in the current
study. However, in the current study the effect of a systematic design
rather than simple random sampling could not be assessed. Neverthe-
less, Gregoire et al. (2011) proposed to use other estimators such as suc-
cessive differences estimators. Ene et al. (in review) tested several such
successive differences estimators and reported that they approximated
the true variances quite well. A similar conclusion was reached by
Nelson et al. (2008) in a simulation study using profiling rather than scanning LiDAR.

The findings of the current study based on data from a real sample supported by simulation results by Ene et al. (2012) clearly indicate
that the 2STMA estimator that has been used in LiDAR sample surveys
in Alaska (Andersen et al., 2009) and Norway (Gobakken et al., 2012;
Gregoire et al., 2011; Nelson et al., 2012) are inappropriate when the flight-lines vary in length. Successive differences estimators will to
some extent compensate for such unequal lengths to the degree that adjacent flight-lines will tend to be more similar in length than flight-lines locate far apart. However, successive differences estimators are
primarily designed to handle systematic sampling in a more approp-
rate way than estimators assuming random sampling. The challenge with unequal cluster sizes is fundamentally a different problem than
what the successive differences estimators are designed to handle. Nev-
evertheless, unequal flight-line lengths will likely be the norm rather than the exception in future LiDAR-assisted sample surveys. It is therefore an
urgent need to develop estimators that can handle such designs prop-
One option could be to develop a model-assisted ratio estimator by taking the individual cluster sizes into account. Such an estimator
would still remain within the design-based and model-assisted ap-
proach. Another option could be to select the primary sampling units
(the LiDAR flight-lines) with probability proportional to size (PPS), i.e., proportional to flight-line length. That was also proposed by
Wulder et al. (2012). However, assuming PPS sampling would not resolve the problems in Hedmark since we employed an equal prob-
ability sampling design, but for future surveys PPS sampling may be
an interesting option.

3.3. Some options for design of LiDAR-assisted sample surveys for REDD

The current study capitalized on an accurately measured sample of
ground plots of the established Norwegian NFI. Most NFIs around the
world position the ground plots with hand-held GPS receivers or simpler
means, which may result in positional errors of a considerable magni-
tude. Positional errors of ground plots may have a detrimental effect on
the estimates for the area of interest (e.g. Gobakken & Næsset, 2009;
McRoberts, 2010). Such errors were avoided in the current study by
using dual-frequency survey-grade receivers (Gobakken et al., 2012) and the influence of GPS errors on the results could therefore be ignored.
Further, the 659 NFI plots were measured during the ninth inventory
cycle of the NFI, which was the third cycle with permanent plots. Experi-
ences indicate that errors such as incorrect tree species classification and crude errors in the field data recording and other non-measurement
errors are continuously removed from a datasets of permanent plots as the re-measurements proceed over time. Thus, even the conventional tree
measurement recordings most likely had a superior quality compared to recently established field surveys. The data from Hedmark therefore offered unique opportunities to study the performance of
LiDAR-assisted sample surveys which currently would be difficult

<table>
<thead>
<tr>
<th>Cover class</th>
<th>Mean</th>
<th>Range</th>
<th>Variance component (%)</th>
<th>N_c</th>
<th>Range</th>
<th>Variance component (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Productive forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>9711</td>
<td>23</td>
<td>30,766</td>
<td>99.0</td>
<td>1.0</td>
<td>18,645</td>
</tr>
<tr>
<td>Medium</td>
<td>23,077</td>
<td>865</td>
<td>67,790</td>
<td>99.1</td>
<td>0.9</td>
<td>45,714</td>
</tr>
<tr>
<td>Low</td>
<td>28,696</td>
<td>1002</td>
<td>64,072</td>
<td>98.1</td>
<td>1.9</td>
<td>50,246</td>
</tr>
<tr>
<td>Young</td>
<td>30,679</td>
<td>954</td>
<td>61,010</td>
<td>96.6</td>
<td>3.4</td>
<td>54,854</td>
</tr>
<tr>
<td>All productive forests</td>
<td>90,880</td>
<td>3007</td>
<td>210,315</td>
<td>97.8</td>
<td>2.2</td>
<td>162,304</td>
</tr>
<tr>
<td><strong>Nonproductive forest and nonforest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonproductive forest</td>
<td>18,767</td>
<td>95</td>
<td>45,577</td>
<td>98.0</td>
<td>2.0</td>
<td>31,763</td>
</tr>
<tr>
<td>Mountain areas</td>
<td>47,433</td>
<td>0</td>
<td>192,173</td>
<td>99.8</td>
<td>0.2</td>
<td>84,406</td>
</tr>
<tr>
<td>Open water</td>
<td>7780</td>
<td>242</td>
<td>27,353</td>
<td>.</td>
<td>.</td>
<td>14,884</td>
</tr>
<tr>
<td>All n.prod. and nonforest</td>
<td>73,980</td>
<td>337</td>
<td>225,346</td>
<td>96.8</td>
<td>3.2</td>
<td>130,595</td>
</tr>
<tr>
<td>All classes</td>
<td>164,860</td>
<td>5600</td>
<td>274,586</td>
<td>96.7</td>
<td>3.3</td>
<td>284,917</td>
</tr>
</tbody>
</table>

The table above shows the contribution (in percent) to the overall estimated variance in two-stage direct estimation (2ST) of the sampling variability in the first (I) and second (II) stages, respectively, for designs assuming E–W and N–S orientation of the first-stage clusters, respectively. The table also displays the mean and range of population elements (N_c) in the first-stage clusters.
in most tropical and REDD-relevant countries where a similar data quality could not be guaranteed or where required data would not be available at all. Although one should indeed aim at repeating the studies reported here under conditions typically found in tropical countries, the lessons learned in this study regarding the choice of appropriate estimators for LiDAR-assisted sample surveys and the potential gains in efficiency that may be achieved should be very valuable and relevant to tropical countries.

The Norwegian NFI — like most other NFIs around the world (see e.g. Tomppo et al., 2010a), follows a simple and systematic design. With permanent sample plots located on a 3 km x 3 km grid the NFI ground survey forced us to adopt a very simple and straightforward design of the LiDAR survey as well, namely a design with parallel and systematically distributed flight-lines with a fixed distance between adjacent lines and thus the same sampling intensity across the entire AOI. Nevertheless, such a simple design is considered to be fairly efficient for a LiDAR campaign since long flight-lines are cost-effective from an operational point of view. Systematic sampling is also generally an attractive design under most circumstances. However, NFIs are historically not designed with remote sensing in mind and existing NFI ground sampling designs may therefore not be “optimal” for applications assuming extensive use of remote sensing data as auxiliary information. Further, NFIs serve multiple purposes among which estimation of biomass and carbon and changes over time are only a few. For REDD carbon monitoring assisted by LiDAR it is for example rather unlikely that it would be efficient with a constant sampling intensity across all different forest types found in the entire AOI.

Several options for improved efficiency may be considered. First, it is likely that pre-stratification can contribute to reduced uncertainties. McRoberts et al. (2006) demonstrated that post-stratification using Landsat data could reduce standard errors of volume estimates, although post-stratification is usually expected to have only marginal or modest impact on the efficiency. Volume and biomass — and thus carbon, are highly correlated. Hence, pre-stratification based on classified satellite remote sensing data from optical satellites like Landsat or synthetic aperture radar into classes assumed to be correlated with the target variables is indeed a viable option. If the objective of the survey is estimation of current biomass and carbon stocks, crude volume/biomass classes or forest classes reflecting differences in biomass may be suitable. Classification of satellite data for stratification purposes will most likely profit from training with local data (supervised classification). In many tropical countries local field data are not available, but even models relating field-observed parameters (like biomass) to remotely sensed observables and fitted with data from another country or even a different continent may still be suited to capture the major geographical distribution and trends of relevance for stratification. One such example is the NFI for Tanzania where the design involving pre-stratification was based on experiences gained through simulations with an artificial population of tree volume of the forest of Tanzania based on Landsat data with individual Landsat pixels as population elements with volume predicted for the individual pixels using regression models for volume fitted with Finnish field data (Tomppo et al., 2010b).

Although pre-stratification may be efficient in LiDAR-assisted sample surveys, it will be challenging to form spatial units of sufficient size to allow the individual flight-lines to be located entirely within the given stratum. From an operational point of view, flight-lines shorter than, say, 100–120 km in length are inefficient as it often takes around 3 min to turn the aircraft and get in position for a new flight-line. It would not take more than around 20 min to collect LiDAR data along a 100 km long strip. Strips longer than 20–25 min flying time is usually not recommended. However, it would not be unrealistic to form units for stratification with a size of, say, 100 km x 100 km for vast areas of for example African savanna.

In areas with abrupt topography and geographically fragmented vegetation one would have to accept shorter flight-lines and lower efficiency of the LiDAR data acquisition. In a recent study in Nepal it was proposed to acquire equal size blocks of LiDAR data (5 km x 10 km), which greatly would facilitate analysis (Anonymous, 2012). The Nepal study did not apply stratification but rather used classification of forest type and proportion forest area for each block derived from Landsat data to select among blocks for subsequent LiDAR sampling and ground sampling according to inclusion probabilities for the blocks determined by weights defined according to forest type and proportion forest area. Each of the selected blocks was covered with “wall-to-wall” LiDAR data. Assuming a swath width of 1000 m, 20% side overlap between adjacent strips to form a block, and 3 min flying time to turn the aircraft, it would take around 30 min to cover a block with size 50 km². If the same resources were used to fly two individual flight lines in sampling mode the area covered would be around 125–130 km². Thus, the sampling fraction would be 2.5 times larger by flying long strips rather than blocks. The block design is therefore most likely inefficient unless the AOI is very fragmented.

An important means to gain efficiency under (pre-) stratified sampling and reduce uncertainty would be to allocate LiDAR strips with unequal sampling intensities to the various strata. In the NFI designed for Tanzania sampling intensities varied by a factor of up to 1:10 for the different strata (Tomppo et al., 2010b). Although the Tanzanian NFI assumed field-sampling only this result should give a clear indication of a potential for improved efficiency by unequal allocation of LiDAR strips to strata even for LiDAR-assisted sample surveys. If the parameter of interest is change in carbon rather than current stocks, the stratification may aim at change over time and may utilize multi-temporal satellite data for classification rather than data from a given point in time. It should be mentioned though that strata will change over time and so will the efficiency. Optimizing for current interests may therefore lead to difficulties later on.

As opposed to (pre-) stratification where the strata are defined in the design-phase and samples are distributed separately within each stratum and independent between strata, post-stratification is conducted after the samples have been allocated. However, post-stratification of LiDAR sample surveys poses a specific problem rarely encountered in post-stratification of field surveys, namely that post-strata boundaries will subdivide the individual flight-lines. Since an individual flight-line (cluster) thus will be composed of several strata, the strata cannot be considered to be independent. This dependency should be accounted for in the uncertainty assessment. Similarly, proper means to account for dependencies among (pre-) strata within individual flight-lines would even increase the flexibility when designing pre-stratified LiDAR surveys (see comments above). Since post-stratification, but certainly pre-stratification, are highly relevant and interesting means to improve efficiency in operational LiDAR-assisted sample surveys (Ene et al., in review), estimators that account for dependencies between strata should be developed, and especially in the context of design-based and model-assisted ratio estimators that also may account for unequal flight-line lengths. It should be noted though that estimators that account for such dependencies already have been developed for model-dependent applications (Stähl et al., 2011).

In post-stratification it is important to pay attention to the field sample size for each post-stratum, especially if the survey also has been (pre-) stratified. If a post-stratification is conducted in a previously pre-stratified survey and the post-strata cut across the pre-strata, then the number of ground plots for a particular combination may be few or some combinations of pre- and post-strata that are present in the population may even suffer from a complete lack of samples (Nasset et al., 2013).

If the analyst is not restricted by an already existing field survey, even the subsequent ground sampling carried out along the LiDAR strips to acquire the secondary sample may be conducted more efficiently when the objective of the survey is estimation of current biomass or carbon stocks or even changes in such stocks. For example, the LiDAR data may be used to predict biomass as a continuous variable for every population element within the LiDAR flight-lines based
on previously fitted models and these predictions may be used to select ground plots proportional to the predicted probabilities. If multi-temporal LiDAR data exist, one may even predict the probability of change or change in biomass and carbon from the LiDAR data using existing models and use these predictions to allocate the field samples proportionally to the predicted probability of change or predicted change in biomass. It has been shown that multi-temporal LiDAR data can be used to predict change in biomass (Bollandsás et al., 2012; Næsset et al., 2013) and even can be used to distinguish between different types of changes (e.g. degradation versus untouched; Næsset et al., 2013). It should be noted that because many types of human activities causing changes in carbon stocks are rare and LiDAR data have a better geographical coverage than field data and yet multi-temporal LiDAR data are highly correlated with changes observed on the ground, using LiDAR data as auxiliary to field data can most likely improve efficiently of change estimation relatively more than it can improve the estimation of current stocks. In the study by Næsset et al. (2013) it was found that in model-assisted estimation of change in biomass using LiDAR data as auxiliary information the relative variance compared to a pure field-based estimate was 1:15 to 1:38 for degraded forest. Degradation was a rare event in the AOI that was analyzed. For untouched forest the gain in precision was more moderate with a relative variance around 1:2 to 1:6. This assumes of course that multi-temporal LiDAR are available, i.e., that the same flight-lines are flown repeatedly over time. An important consideration when designing LiDAR-assisted sample surveys is therefore whether the same flight-lines should be flown in subsequent surveys.

Finally, it should be mentioned that “optimal” characteristics of the individual field plots may differ somewhat between a conventional field survey and a field sample collected as part of an overall design involving auxiliary data from LiDAR. Of particular importance is the size and shape of the individual plots. Experience with different plot sizes in LiDAR-assisted forest inventory is limited, although increasing plot size will tend to improve accuracy (Frazer et al., 2011; Golakken & Næsset, 2008). An interesting observation was made by Næsset et al. (2011, p. 3611) though. They compared uncertainties of biomass estimates obtained using LiDAR data as auxiliary to field data in a model-assisted estimation with the uncertainty of a pure field-based estimate and noted that when the plot size increased the precision improved relatively more for the model-assisted estimate. This relative improvement of precision was attributed to smaller relative impact of edge effects and GPS positioning errors with larger plots. Edge effects are caused by mismatch between the trees that are defined to be inside the plot on the ground and thus measured in the field survey and the tree crowns that are measured from the air. Trees on the ground close to the edge of the plot may partly have the crowns outside the plot while trees with the stems located outside the plot may partly have the crowns inside the plot. An interesting illustration of this phenomenon is given in Næsset et al. (2013). Thus, there is a potential for improved precision with increasing plot size and appropriate plot sizes should be determined with due attention to how they are going to be used since (2013). Thus, there is a potential for improved precision with increasing plot size and appropriate plot sizes should be determined with due attention to how they are going to be used since (2013).

4. Conclusions

To conclude, this study has shown that LiDAR-assisted sample surveys can improve precision substantially compared to pure field-based surveys. The empirical results for the particular population under study suggest a potential reduction in standard error of around 40–60% compared to that obtained in a pure field survey. Unequal flight-line lengths are often practical from an operational perspective but recently proposed variance estimators do not account for such unequal line lengths and will thus tend to overestimate the true but unknown sampling variability and cannot be recommended for use in LiDAR-assisted surveys if the line lengths deviate much. We propose to develop new estimators accounting for unequal lengths. The current LiDAR survey was restricted by the simple and systematic design of the existing Norwegian NFI. Experience from design of field-based surveys and previous research with optical satellite remote sensing indicate a potential for substantial gain in precision also for LiDAR-assisted surveys by (pre-) stratified designs and even post-stratified estimation which would be of great value for surveys in tropical and REDD-relevant countries where LiDAR surveys to a less extent would be restricted by existing field surveys. New design-based and model-assisted two-stage estimators that account for dependencies between pre- and post-strata when an individual flight-line is composed of several strata should be developed. That will allow for greater flexibility in design and estimation in future LiDAR-assisted surveys in tropical countries. Appropriate plot sizes for LiDAR-assisted surveys will tend to be somewhat larger than in pure field surveys, and use of circular plots whenever feasible will simplify GPS positioning and co-registration with remotely sensed data and reduce the impact of edge effects compared to rectangular plots.

Acknowledgments

This research has been funded by the Research Council of Norway (project #166482/i10 and #184636/S30). The authors would like to thank the Norwegian National Forest Inventory for collection of field data and Blom Geomatics, Norway, for collection and processing of the airborne LiDAR data.

References


