Remote Sensing of Environment 133 (2013) 210–224

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys

Liviu Theodor Ene [a,*], Erik Næsset [a], Terje Gobakken [a], Timothy G. Gregoire [b], Göran Ståhl [c], Sören Holm [c]

[a] Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway
[b] School of Forestry and Environmental Studies, 360 Prospect St., Yale University, New Haven, CT 06511-2189, USA
[c] Dept. of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden

ARTICLE INFO

Article history:
Received 2 October 2012
Received in revised form 3 February 2013
Accepted 7 February 2013
Available online 17 March 2013

Keywords:
Forest inventory
Airborne laser scanning
Design-based inference
Model-based inference
Variance estimation
Monte-Carlo simulations
Cost-efficiency

ABSTRACT

Auxiliary information provided by airborne laser scanners (ALS) is expected to increase the accuracy of biomass estimation in large-scale forest surveys. Because acquisition of "wall-to-wall" ALS data over large areas is not economically feasible, a systematic sampling approach using ALS as a strip sampling tool was used to supplement a conventional field-based inventory in a large-scale biomass survey in Hedmark County (HC), Norway. However, the complexities of these surveys render prohibitive the analytical determination of the properties of the resulting estimators and of the estimators of their sampling variances. To overcome the problem, the statistical properties of the estimators were empirically investigated in this paper using simulated sampling from an artificial population. Through this approach, estimators with desirable properties can be identified and used for inference in real applications. By combining biomass estimates from Norwegian National Forest Inventory plots in HC, ALS measurements and Landsat 5 TM imagery, an artificial population at the scale of HC was created. Using this artificial population as "ground-truth", we demonstrate how simulated sampling can be used for assessing the statistical properties of regression estimators and of their variance estimators under two-phase post-stratified systematic sampling (*SYS*) and simple random sampling without replacement (*SRSwoR*) designs, considering design- and model-dependent inferential frameworks. The results were assessed using a purely ground-based systematic design with a Horvitz–Thompson (*HT*) estimator as benchmark. The real overall precision of the ALS-aided systematic survey was nearly five times overestimated when using the design-based variance estimators developed for SRSwoR, while under model-dependent inference the overestimation of the real standard errors was around 40%. Compared to ground-based inventory, the estimated standard errors of the systematic ALS survey doubled while in reality the standard errors were 55% lower. Using successive differences variance estimators greatly improved the precision of the systematic ALS-aided survey and produced valid 95% confidence intervals under the design-based inference. The most satisfactory results for the ALS-aided survey in terms of analytical variances occurred under design-based inference with successive difference variance estimator, closely followed by the model-dependent estimators. Using simulations, the cost efficiency of the ground based and ALS-aided surveys was assessed by evaluating accuracy against inventory cost for various sampling intensities. The results indicated that the ALS-aided surveys can be a cost-efficient alternative to traditional field inventories.

1. Introduction

With increasing concerns regarding global climate change, accurate estimation of forest carbon pools using remotely sensed auxiliary information has become a very active research field. Forest ecosystems are known to store large amounts of carbon (IPCC, 2000), and one of the most dynamic and largest carbon pools is the aboveground biomass (AGB) pool of living trees (Fahey et al., 2010). In Norway, the Norwegian National Forest Inventory (NFI) (Tomter et al., 2010) has undertaken reporting of carbon stock changes from the Land Use, Land Use Change and Forestry (LULUCF) sector (IPCC, 2006) required by the Kyoto Protocol. In tropical countries, an emerging interest for large-scale carbon inventories is driven by the initiatives to reduce emissions from deforestation and degradation (REDD) (Asner, 2009; Gibbs et al., 2007).

However, terrestrial carbon inventory systems may not be economically feasible for large area surveys, and may not provide reliable local estimates, for instance at sub-regional and administrative unit level or by land-use or cover classes (Fahey et al., 2010; Gonzales et al., 2010). Combining auxiliary information provided by remote sensing systems with terrestrial surveys has the potential to enhance the

* Corresponding author. Tel.: +47 64965766; fax: +47 64968890.
E-mail addresses: Liviu.Ene@umb.no (L.T. Ene), Erik.Naesset@umb.no (E. Næsset), Terje.Gobakken@umb.no (T. Gobakken), Timothy.Gregoire@yale.edu (T.G. Gregoire), Goran.Stahl@slu.se (G. Ståhl).

0034-4257/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2013.02.002

precision of forest carbon estimation (Gonzales et al., 2010; Næsset et al., 2011). Having the ability to describe the three-dimensional canopy structure, Airborne Light Detection and Ranging (LiDAR) is a promising remote sensing tool for enhancing biomass inventories in various forest types (Asner et al., 2009; Boudreau et al., 2008; Drake et al., 2003; Lefsky et al., 2002; Næsset, 2011; Næsset & Gobakken, 2008; Næsset et al., 2011; Nelson et al., 2007, 2004). However, "wall-to-wall" LiDAR surveys are still expensive for large areas, and economically feasible emerging technologies like the Interferometric Synthetic Aperture Radar (InSAR) (Gama et al., 2010; Næsset et al., 2011; Solberg et al., 2010a,b) or fusion of LiDAR and InSAR (Sun et al., 2011) have been tested as an alternative for LiDAR-based AGB estimation.

An alternative to the expensive "wall-to-wall" surveys is the use of LiDAR measurements within a sampling framework. Parker and Evans (2004), Andersen and Breidenbach (2007), Corona and Fattorini (2008), and Stephens et al. (2012) discussed double-sampling applications using airborne laser scanning (ALS) measurements on subsampled field plots. Line-intersect sampling for AGB and forest volume estimation using profiling airborne lasers such the Portable Airborne Laser System (PALS; Nelson et al., 2003) was tested in areas up to 5000 km^2 in size (Nelson et al., 2008, 2004). Three-phase surveys using spaceborne laser, PALS and field plots as the first, second, and third phase samples, respectively, were described by Boudreau et al. (2008) and Nelson et al. (2009).

An alternative design was suggested by Næsset (2005) and Gobakken et al. (2006), considering ALS as a strip sampling tool, where the LiDAR measurements are collected along corridors of several hundred meters wide aligned with the ground plot network. Under this design, the LiDAR data itself becomes a sample, introducing additional uncertainty while reducing the data acquisition costs. Andersen et al. (2009) applied this design in a two-stage survey for biomass estimation in Kenai Peninsula of Alaska, USA. In Norway, this design was tested in a large-scale biomass survey covering nearly 30,000 km^2 in Hedmark County (HC), in the period 2005–2010, using ALS, PALS, Norwegian NFI data and supplementary ground plots (Næsset et al., 2009).

The estimation in the HC survey was performed using regression estimators under two inferential frameworks: (1) two-stage, model-assisted (DBMA) by Gregoire et al. (2011; Gobakken et al., 2012), and (2) two-phase, model-dependent (MD) by Ståhl et al. (2011; Gobakken et al., 2012). Gregoire (1998), Schreuder et al. (2001), and Kangas (2006) present more details regarding the design- and model-based inferential frameworks.

The results reported in the HC project (Gobakken et al., 2012; Gregoire et al., 2011; Ståhl et al., 2011) indicated little differences between the estimated precisions of LiDAR-based estimates and those obtained using the pure ground-based sampling, which may indicate the overestimation of sampling variance in complex LiDAR-based surveys. As an alternative for identifying an efficient sampling strategy, Ene et al. (2012) used a Monte-Carlo approach for assessing the statistical properties of the DBMA and MD estimators used in HC survey.

The results presented by Ene et al. (2012) indicated that violating the assumptions the variance estimators were built upon (e.g., using systematic sampling (*SYS*) instead of simple random sampling without replacement (*SRSwoR*)) greatly overestimated the uncertainty of the AGB estimates. Compared to the field-based survey, the simulation results indicated that the precision of the ALS-based estimation improved by 59%, while the estimated standard errors indicated the opposite, i.e. that the standard errors were 1.8 times higher. Moreover, the estimated standard errors due to the first-phase sampling dominated the total estimated sampling error. This suggests adopting variance estimators specifically constructed for dealing with systematic sampling, like that based on successive differences (Cochran, 1977; Wolter, 2007), as suggested by Gregoire et al. (2011) and Ståhl et al. (2011).

In the current paper, the simulation study performed by Ene et al. (2012) for unstratified designs is further developed considering post-stratification and testing alternative variance estimators for systematic sampling. Thus, the objectives of this study were:

1. to assess the bias of the variance estimators proposed by Gregoire et al. (2011) and Ståhl et al. (2011) for post-stratified AGB estimation under *SRSwoR* and *SYS* designs;
2. to assess the performance of successive differences estimators under two-phase systematic sampling designs;
3. to assess the efficiency of ALS-aided and pure ground-based AGB surveys;
4. to assess the design effects of the two-phase systematic and simple random sampling designs, and
5. to evaluate the relative costs of the ALS aided survey and field inventory under systematic sampling.

The proposed methodologies are adopted in a case study involving post-stratified AGB estimation, where the administrative units represented the post-strata.

2. Material

The study area was Hedmark County located in south eastern Norway (Fig. 1). The county has a land area of 27,399.72 km^2 and contains four main administrative units (AU). HC is rather heterogeneous with regard to forest types and geomorphology. The dominant tree species present in HC are Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.).

The material comprises three datasets: (1) field inventory data, (2) remote sensing data (ALS and satellite imagery), and (3) cartographical products in the form of a digital terrain model (DTM) as raster data and land-use maps. For a more detailed presentation of the datasets, see Ene et al. (2012) and Gobakken et al. (2012).

2.1. Satellite imagery and map data

An image mosaic providing full coverage of HC was build using the bands 1 to 5 and 7 of three nearly cloud-free Landsat 5 TM images. The spatial resolution of the mosaic was 15.81 m such that the area covered by a pixel was 250 m^2.

The map data consisted of a 25 m spatial resolution DTM produced by the Norwegian Mapping Authority from the official topographic map series, and *AR5* land use maps (Bjørdal & Bjørkelo, 2006). Using the DTM and the land-use map, a forest mask was created for the entire HC. The DTM was used for delineating the coniferous tree line in HC (Ørka et al., 2012), and the land-use maps were used for eliminating water and other types of non-forested areas. A forest vegetation mask representing about 86% of the tessellated area of HC was created for the developed areas and for the regions between 850 and 1150 m a.s.l.

2.2. Field data

The field observations were acquired from the permanent plots of the Norwegian National Forest Inventory. The Norwegian NFI is a continuous forest inventory system having the sampling units represented by fixed-area circular plots of 250 m^2 displaced in 3×3 km grid covering the entire country, except the areas above the coniferous tree line where the grid spacing is 3×9 km. The grid axes are oriented to the north–south and east–west directions. Due to the Latin square design used for deciding the revisiting order of the field plots across the country (Tomter et al., 2010), approximately 20% of the 2309 NFI plots located in HC are measured every year. Totally, 1483 plots (including NFI plots and additional plots subjectively located in developed and mountain areas) were measured during three field campaigns in the period