Remote Sensing of Environment 158 (2015) 431–440

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information

Svetlana Saarela [a,*], Anton Grafström [b], Göran Ståhl [b], Annika Kangas [a], Markus Holopainen [a], Sakari Tuominen [c], Karin Nordkvist [b], Juha Hyyppä [d]

[a] Department of Forest Sciences, University of Helsinki, PO Box 27, FI-00014 Helsinki, Finland
[b] Department of Forest Resource Management, Swedish University of Agricultural Sciences, SLU Skogsmarksgränd, SE-901 83 Umeå, Sweden
[c] Finnish Forest Research Institute, PO Box 18, FI-01301 Vantaa, Finland
[d] Finnish Geodetic Institute, PO Box 15, FI-02431 Masala, Finland

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 15 November 2014
Accepted 17 November 2014
Available online 11 December 2014

Keywords:
Airborne LiDAR
Landsat
Model-assisted estimator
Probability sampling
REDD monitoring

ABSTRACT

Airborne Light Detection and Ranging (LiDAR) and Landsat data were evaluated as auxiliary information with the intent to increase the precision of growing stock volume estimates in field-based forest inventories. The aim of the study was to efficiently utilize both wall-to-wall Landsat data and a sample of LiDAR data using model-assisted estimation. Variables derived from the Landsat 7 ETM + satellite image were spectral values of blue, green, red, near infra-red (IR), and two shortwave IR (SWIR) bands. From the LiDAR data twenty-six height and density based metrics were extracted. Field plots were measured according to a design similar to the 10th Finnish National Forest Inventory, although with an increased number of plots per area unit. The study was performed in a 30000 ha area of Kuortane, Western Finland. Three regression models based on different combinations of auxiliary data were developed, analysed, and applied in the model-assisted estimators. Our results show that adding auxiliary Landsat and LiDAR data improves estimates of growing stock volume. Very precise results were obtained for the case where wall-to-wall Landsat data, LiDAR strip samples, and field plots were combined; for simple random sampling of LiDAR strips the relative standard errors (RSE) were in the range of 1–4%, depending on the size of the sample. With only LiDAR and field data the RSEs ranged from 4% to 25%. We also showed that probability-proportional-to-size sampling of LiDAR strips (utilizing predicted volume from Landsat data as the size variable) led to more precise results than simple random sampling.

1. Introduction

Forest resources are required for an increasing number of purposes globally, including wood- and fibre-based raw materials, maintenance of biodiversity, and mitigation of climate change (Mery et al., 2005). As a consequence, the demands for information from forests are steadily increasing (Kangas & Maltamo, 2006; UNECE and FAO, 2011). National forest inventories (NFIs) have been established for a long time in many countries (e.g. Tomppo et al., 2010). Normally, they are based on statistical samples of field plots (McRoberts et al., 2009, 2010; Woodall et al., 2009) as a means for ensuring trustworthy information, i.e. information derived from estimators that are unbiased and have high precision.

Field-based forest inventories have many advantages. However, they become expensive when large sample size is required to reach the needed levels of precision. Furthermore, sparse road networks or other conditions in a country may prevent easy access to the plots. Also, NFI information from field plots alone often leads to imprecise estimates for small regions within a country. This has stimulated the development of solutions where field plots and remotely sensed data are combined in order to provide the required information (Holmström et al., 2001; Maltamo et al., 2007; Næsset, 2004).

Lately, the REDD + mechanism (reducing emissions from deforestation and forest degradation; Brockhaus (2009)), which has been developed under the United Nations' Framework Convention on Climate Change, has led to an even stronger focus on forest information and NFIs, and on how to utilize remote sensing within NFIs, especially in countries with poor infrastructure conditions. Several approaches based on remote sensing have been developed and demonstrated , (e.g. Gobakken et al., 2012; Næsset et al., 2006; Nelson et al., 2009). However, inventories that make use of auxiliary information from remote sensing are not only relevant for developing countries and REDD + (e.g. Asner, 2009; Saatchi et al., 2011), but also for remote areas in developed countries, such as Siberia and Alaska (e.g. Andersen et al., 2009; Ene et al., 2012; Nelson et al., 2009). Further, in countries with well-established field-based NFIs, sample-based combinations of

* Corresponding author at: Latokartanonkaari 7, 00014 University of Helsinki, Finland. Tel.: +358 442177795; fax: +358 919158100.
E-mail address: svetlana.saarela@helsinki.fi (S. Saarela).

http://dx.doi.org/10.1016/j.rse.2014.11.020
0034-4257/

field and remotely sensed data may offer new possibilities to make inventories cost-efficient.

The problems involved in reaching good inventory solutions include variable and sometimes limited information in remotely sensed data, the need to combine remote sensing with field information in order to obtain reliable results, the lack of adequate field samples, the need to apply advanced statistical methods, and the challenge to make the solutions straightforward enough so that they can be easily operated in practice.

Regarding the first issue, it is well known that remotely sensed data only contain auxiliary information about a limited number of all the variables that are typically addressed in NFIs. For example, land-use classification can seldom be based on image data alone (Anderson, 1976; Lo & Choi, 2004; Shalaby & Tateishi, 2007). However, for some important variables – for example tree biomass and volume – several remote sensing techniques such as LiDAR and RADAR have great potential (e.g. Wu, 1987; Hyyppä & Hallikainen, 1996; Boudreau et al., 2008; Asner et al., 2011; Næsset, 2011; Bollandsås et al., 2013). Thus, the focus on biomass and carbon in many emerging inventories makes remote-sensing-based solutions potentially very useful.

LiDAR data are known to provide auxiliary data that are highly correlated with growing stock volume, biomass and aboveground carbon in forests (Hyyppä & Inkinen, 1999; Hyyppä et al., 2008; Næsset, 1997; Næsset, 2009, 2011; Næsset & Gobakken, 2008; Nelson et al., 1988; Stephens et al., 2012). In many applications, LiDAR data have been acquired wall-to-wall over the target forest areas and stand-level estimates have been derived either based on the area method (Næsset, 2011) or based on the identification of individual trees (Hyyppä et al., 2001). For applications over large areas, such as countries, the acquisition of LiDAR data is prohibitively expensive; however, the data acquisition can be carried out as part of a sampling scheme to improve the precision of estimates. For example, Nelson et al. (2004) used a profiling LiDAR to estimate the forest resources of Delaware and Andersen et al. (2009) used data from an airborne laser scanner to estimate forest resources within a region of Alaska.

The statistical inference may be either model-based (e.g. McRoberts, 2010; Ståhl et al., 2011) or design-based (e.g. Gregoire et al., 2011). In the design-based approach, model-assisted estimators (Gregoire et al., 2011; Næsset et al., 2013) are typically used based on data from a probability sample; the deviations between the model predictions and the field reference data are calculated and used to correct the model predictions. Main advantages of this approach are that all the attractive properties of design-based inference can be utilized, while at the same time, the LiDAR data and the model can improve the precision of estimates substantially. Drawbacks include that a strict probability sample from the entire population must be acquired and that mismatches in geopositioning between field plots and remotely sensed data may have a negative effect on the accuracy of estimators.

Several studies have been conducted where LiDAR samples and field samples have been combined utilizing two-phase sampling and model-assisted estimation (e.g. Gregoire et al., 2011; Næsset et al., 2011); multi-spectral satellite data have sometimes been used to stratify the target population in these studies. Results have highlighted the usefulness of LiDAR as auxiliary data, although in some cases the gain over traditional field sampling has been modest (e.g. Gobakken et al., 2012). However, so far relatively few studies have been performed where several sources of remotely sensed predictors have been utilized in connection with model-assisted estimators. Examples include Andersen et al. (2012), who applied LiDAR and satellite data in a post-stratification approach utilizing multilevel sampling, and Strunk et al. (2014), who utilized several sources of remotely sensed data and compared linear regression and k nearest neighbours (kNN) methods to link field reference data, LiDAR and Landsat data.

In developing this type of inventory the choice of sampling strategy is important. It is well known (e.g Särndal et al., 1992) that auxiliary information can be utilized both for improving the design, i.e. how the sample is selected, and for improving the estimators, i.e. how the target quantities are computed once data have been acquired. Most studies so far have utilized fairly straightforward sampling designs, such as simple random or systematic sampling of LiDAR strips that traverse the entire study area (e.g. Andersen et al., 2009; Gobakken et al., 2012). Auxiliary data have been used to improve the estimators. However, as shown by Grafström et al. (2014), the potential gain from choosing an appropriate sampling design may be substantial and needs to be further evaluated.

The objective of this study was to explore how three sources of information that would typically be available in connection with LiDAR sample surveys could be combined utilizing model-assisted estimation. Our data were wall-to-wall Landsat data, strip samples of LiDAR data, and field plots. We compare different cases of auxiliary data usage, show how estimates as well as uncertainty estimates can be derived, evaluate the effect of different sampling designs incorporating LiDAR, and discuss the advantages and disadvantages of the proposed sampling strategies. A novel feature of the study was that a strategy based on probability-proportional-to-size sampling was evaluated for the selection of LiDAR strips. The study was performed in the Kuortane area in the boreal forests of western Finland.

2. Materials

2.1. Study area

The Kuortane study area is located in western Finland in the southern Ostrobothnia region (see Fig. 1). It is mainly covered by middle-aged Scots pine boreal forest in the Suomenselkä watershed area. Norway spruce and deciduous trees, mainly birches, usually occur as mixtures. The landscape is composed of forests on mineral soils, peatlands drained for forestry, open mires, and agricultural fields at lower elevations. The terrain depressions are covered by lakes.

Non-forest areas – about one third of the total area – were masked out using digital map data (Tomppo et al., 2008). The area was tessellated into 16×16 m grid cells for which Landsat and LiDAR data were acquired. Our population was 818 017 grid cells corresponding to 20 942 ha of forest.

2.2. Data

The material comprised three datasets: field data (Section 2.2.1), airborne LiDAR data (Section 2.2.2), and Landsat 7 ETM+ data (Section 2.2.3).

2.2.1. Field data

The field data were sampled using a modification of the design of the Finnish NFI (Tomppo et al., 2008), where the sampling density was significantly increased and the number of measured variables was somewhat reduced. The NFI is a sample-based inventory system that covers all land-use classes and ownership categories in Finland. The aim of the NFI is to produce reliable information on forest resources at national and regional level. The NFI is based on statistical sampling. The sampling design is systematic cluster sampling. Here, the plots were clustered into rectangular clusters of 18 plots, with 200 m distance between plots within clusters. The distance between clusters was 3 500 m.

Each circular field plot has a radius of 9 m, so that the size of a plot area corresponds to the grid cell size of the tessellated area. Trees with a diameter at breast height (DBH) larger than 5 cm were measured as tally trees. DBH, tree storey class and tree species were recorded for each tally tree. Height was measured for one sample tree of each species and storey class per plot. The height measurements were used to calibrate the height estimates of the tree species-specific height models of Veltheim (1987). Volume models (Laasasenaho, 1982) were used to estimate tree volumes. The tree-level volumes were then transformed to volumes per hectare for each plot. The field plot locations were