Remote Sensing of Environment 172 (2016) 101–108

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Effects of positional errors in model-assisted and model-based estimation of growing stock volume

Svetlana Saarela [a,*], Sebastian Schnell [b], Sakari Tuominen [c], András Balázs [c], Juha Hyyppä [d], Anton Grafström [b], Göran Ståhl [b]

[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-90183 Umeå, Sweden
[c] Natural Resources Institute Finland (Luke), PO Box 18, FI-01301 Vantaa, Finland
[d] National Land Survey of Finland, Finnish Geospatial Research Institute (FGI), PO Box 15, FI-02431 Masala, Finland

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form 25 October 2015
Accepted 10 November 2015
Available online 14 November 2015

Keywords:
Design-based inference
Forest inventory
Location errors
Model-based inference
Remote sensing
Sampling simulation

ABSTRACT

Positional errors may cause problems when field and remotely sensed data are combined in connection with forest surveys. In this study we evaluated the effects of such errors on statistical estimates of growing stock volume using model-assisted and model-based estimation. With model-assisted estimation, positional errors affect the model parameter estimates for the models that are used as part of the estimation framework. In addition, positional errors affect the estimators, since the deviations between model predictions and field measurements are often larger than they would have been without positional errors. Using model-based estimation positional errors affect the model parameter estimates and thus the estimators. We compared the effects of positional errors in model-assisted and model-based estimation through Monte Carlo sampling simulation in a simulated study area resembling the forest conditions in Kuortane, western Finland. The forest population was created using a copula modelling approach based on field, Landsat and LiDAR data. We found that positional errors led to slightly biased estimators, and estimators with larger variances compared to the cases where data were perfectly geo-located. The relative increase of the variances of the estimators was of equal magnitude for model-assisted and model-based estimation, when models were developed and applied to data with geopositional errors. Further, the variance estimators were always more precise for the model-based estimators compared to the model-assisted estimators. When the models were developed based on perfectly geo-located data but applied to data with positional errors, model-based estimation was superior to model-assisted estimation.

1. Introduction

Forest surveys based on combinations of field and remotely sensed (RS) data are becoming increasingly important (e.g., McRoberts, Tomppo, & Næsset, 2010). Through this combination, forest information can be obtained both in terms of maps and statistical estimates of means and totals of target variables (Tomppo, Olsson, et al., 2008). Also, RS data typically are much less expensive than field data per area unit and thus cost-efficient sampling designs can be developed that require fewer field plots compared to traditional forest inventories. In several countries, infrastructure in terms of road networks is poorly developed and surveys that rely largely on RS data may be the only alternative; e.g., this is the case for many countries in the tropics that are currently preparing for implementing the REDD+ mechanism (reducing emissions from deforestation and forest degradation) under the United Nations' framework convention on climate change (United Nations, Kyoto, 1998; Cienciala et al., 2008).

Regarding statistical estimates, different inferential frameworks are available for the combination of field and RS data. A well-known approach is to use RS data only for stratification or post-stratification (e.g., McRoberts, Wendt, Nelson, & Hansen, 2002; Nilsson et al., 2003). More advanced, and currently rather intensively studied, methods include model-assisted and model-based estimation (e.g., Breidenbach & Astrup, 2012; Gregoire et al., 2011; Næsset, Bollandsås, Gobakken, Gregoire, & Ståhl, 2013; Opsomer, Breidt, Moisen, & Kauermann, 2007; Saarela, Grafström, et al., 2015; Ståhl et al., 2011). With model-assisted estimation (Saarela, Grafström, et al., 2015), regression models are developed based on data from plots where both field and RS data are available. The models are applied to all areas for which RS data are available; this may be a complete enumeration or a sample of the target area (e.g., Saarela, Grafström, et al., 2015). A total value of the target variable based on the model predictions is estimated. Subsequently this value is corrected by a sample-based estimate of the total of the deviations between values based on measurements in the field and the model predictions. Thus, model-assisted estimation requires a probability sample of field plots to be available from the entire target area. The theory for model-assisted estimation has been developed,

* Corresponding author at: University of Helsinki, Latokartanonkaari 7, 00014, Finland.
E-mail address: svetlana.saarela@helsinki.fi (S. Saarela).

http://dx.doi.org/10.1016/j.rse.2015.11.002
0034-4257/

largely, by Särndal, Swensson, and Wretman (1992), although standard regression estimators have been available for a long time (e.g., Cochran, 1977). Applications of model-assisted estimation for assessing forest resources include e.g., Opsomer et al. (2007), Gregoire et al. (2011), Breidenbach and Astrup (2012), Strunk, Temesgen, Andersen, and Packalén (2014) and Saarela, Grafström, et al. (2015).

With model-based estimation, models are developed using data from plots where both field and RS data are available (McRoberts, 2006; McRoberts, Næsset, & Gobakken, 2013). These models are then applied to the entire target area (Ståhl et al., 2011). Totals are obtained by summing the predictions and mean values by dividing the totals with the known total area. The theory for model-based estimation dates back to Matérn (1960), and important contributions have been made by Royall (1971) and Cassel, Särndal, and Wretman (1977). Applications in forest inventories are described by Gregoire (1998), Magnussen, McRoberts, and Tomppo (2009), McRoberts et al. (2013), Ståhl et al. (2011) and Ståhl, Heikkinen, Petersson, Repola, and Holm (2014).

In some cases RS data are only available from a sample of the target area; in this case the inferential framework will be a hybrid of design-based and model-based estimation (Corona, Fattorini, Franceschi, Scrinzi, & Torresan, 2014; Ståhl et al., 2011). Contrary to model-assisted estimation, model-based estimation does not require a probability sample of field plots to be available from the target area. Field plots for the model parameter estimation may be purposively selected.

The tradition of conducting sample-based field inventories, such as national forest inventories (NFIs), dates back about a hundred years in several countries (Tomppo et al., 2010). The tradition of conducting sample-based forest inventories based on combinations of field and RS data has a much shorter history (Tomppo, Haakana, Katila, Peräsaari, 2008), but are available in countries such as in Switzerland (Brassel & Lischke, 2001), Italy (Tabacchi et al., 2007) and the U.S.A. (McRoberts et al., 2010). Further, several issues remain to be studied and resolved, e.g. how estimates should efficiently be broken down on different domains of study and what positional accuracies are required in the determination of the location of field and RS data in order to assure accurate estimators of target population parameters. The latter issue has been studied to some extent by Reese, Granqvist-Pahlén, Egberth, Nilsson, and Olsson (2005), and it has been shown that positional accuracies in the order of 5 m can seriously affect plot level predictions based on RS data. Further, Zhang et al. (2013) investigated the uncertainties of mapping aboveground forest carbon as a result of geo-location errors of sample plots and Landsat TM data. The authors showed a clear positive correlation between regression model goodness of fit and location error distance. Similar results were reported by Wang et al. (2011). However, the connection between geo-location errors and model-assisted and model-based estimation of forest resources has only been studied by McRoberts (2010), where effects of rectification and Global Positioning System (GPS) errors on satellite image-based inference for proportion forest area was analysed, and by McRoberts et al. (2005), where effect of perturbing and swapping inventory pots locations was estimated. For applying these techniques in practice this is an important issue, since using accurate positioning equipment in forest may be difficult and time consuming in order to achieve adequate accuracy in a cost-efficient way, because there are several factors, such as dense canopy cover, that negatively affect the Global Navigation Satellite System (GNSS) signals. Further, the impacts of positional errors are likely to be different between different inferential frameworks, since in model-assisted estimation co-location of field plots and RS data are needed both for model estimation and application, whereas in model-based inference co-location is needed only for model development. Thus, this issue might be important to consider when choosing what inferential framework should be applied in a forest survey.

The objective of this study was to assess the effects of positional errors in field and RS data in connection with model-assisted and model-based estimation. In particular, we evaluated any differences in the effects of positional errors between the two inferential frameworks. Our study focused on standard errors of growing stock volume (GSV) estimates.

2. Material and methods

2.1. Simulated population

The study was conducted through sampling simulation within a simulated forest. The simulated forest was constructed following the method developed by Ene et al. (2012) using data collected from the Kuortane region in western Finland. The area was chosen for a pilot research project using Airborne Light Detection And Ranging (LiDAR) data in forest inventories. The LiDAR data acquisition was done in July 2006 using an Optech 3100 laser scanning system. The average flying altitude above terrain was 2000 m. The mean footprint diameter was 60 cm and the average point density was 0.64 m^{-2}. Altogether 19 north–south oriented flight lines were flown using a side overlap of about 20%.

In this study area, a modified Finnish NFI measuring system was employed for the field survey. The sampling density was increased, e.g., each cluster had 18 plots instead of the usual 9, and the density of clusters was almost doubled. The distance between plots in a cluster was 200 m, between clusters was 3500 m. Circular NFI plots with a fixed radius of 9 m were used. Every seventh tally tree was measured as a sample tree. The diameter at breast height was measured for all trees larger than 50 mm. Tree heights were estimated using tree species specific height models by Veltheim (1987), and the estimated heights were calibrated using height measurements from the sample trees. For each tree with a diameter larger than 50 mm, tree-level volumes were estimated (Laasasenaho, 1982). The tree-level volumes were transformed to volumes per hectare for each plot (Tomppo, Haakana, et al., 2008). Plot locations were assessed with Trimble ProXH with an RMSE of less than 1 m. A total of 441 field plots were available for forest areas. Areas of other land use classes were masked out using digital land-use maps.

Based on findings in previous studies using the Kuortane dataset (Saarela, Grafström, et al., 2015; Saarela, Schnell, et al., 2015), the features of LiDAR data used as auxiliary information for each plot were: maximum height ($h_{\max}$), 80th percentiles of the height distribution (h_{80}), canopy relief ratio (CRR), and percentage of first returns above 2 m (p_{veg}) as crown cover estimate. These features were computed from the laser scanning point cloud using the FUSION software (McGaughey, 2012) and the Orientation and Processing of Airborne Laser Scanning data (OPALS) software (Pfeifer, Mandlburger, Otepka, & Karel, 2014). In addition to LiDAR data, we downloaded Landsat orthorectified (L1T) imagery data from U.S. Geological Survey (2014). The Landsat 7 Enhanced Thematic Mapper (ETM+) multispectral data were acquired in June 2006 (path 90 and row 16). For each field plot digital numbers of spectral values corresponding to the green (B20), red (B30) and shortwave infra-red (B50) Landsat bands were computed, using the nearest neighbour re-sampling method in the ArcGIS software (ESRI, 2011).

The same simulated population as in Saarela, Schnell, et al. (2015), resembling the Kuortane study area forest conditions in western Finland, was used in this study. The population was created using a multivariate probability distribution copula technique (Ene et al., 2012; Nelsen, 2006) based only on plots with non-zero growing stock values. Our study population consists of 818,016 grid cells of 16 m × 16 m size, located in the land-use category forest. For each grid cell, values of Landsat spectral values, LiDAR metrics, and GSV were simulated. In order to perform our analysis of errors related to geopositional mismatching in cases when a boundary cell was randomly selected through the Monte Carlo simulation approach, we created a buffer zone outside the boundary grid cells (see Fig. 1). The buffer zone assured that all boundary grid cells have eight neighbouring grid cells. Grid cells of the buffer zone were taken as a random sample from the