Remote Sensing of Environment 154 (2014) 226–233

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Estimation for inaccessible and non-sampled forest areas using model-based inference and remotely sensed auxiliary information

Ronald E. McRoberts [a], Erik Næsset [b], Terje Gobakken [b]

[a] Northern Research Station, U.S. Forest Service, 1992 Folwell Avenue, Saint Paul, MN USA
[b] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form 14 August 2014
Accepted 15 August 2014
Available online 18 September 2014

Keywords:
Landsat
Lidar
Precision

ABSTRACT

For remote and inaccessible forest regions, lack of sufficient or possibly any sample data inhibits estimation and construction of confidence intervals for population parameters using familiar probability- or design-based inferential methods. Although maps based on remotely sensed data may provide information on the distribution of resources, map-based estimates are subject to classification and prediction error, and map accuracy measures do not directly inform the uncertainty of the estimates. Model-based inference does not require probability samples and when used with synthetic estimation can circumvent small or no-sample difficulties associated with probability-based inference. The study focused on estimating proportion forest area using Landsat data for a study area in Minnesota, USA, and aboveground biomass using airborne laser scanning data for a study area in Hedmark County, Norway. For both study areas, model-based inference was used to estimate the components necessary for constructing confidence intervals for population means for non-sampled areas. The estimates were compared to simple random sampling, model-assisted, and model-based estimates that would have been obtained if the areas had been sampled. All estimates were within two simple random sampling standard errors of each other, thereby illustrating the utility of model-based inference for non-sampled areas.

Published by Elsevier Inc.

1. Introduction

1.1. Background and motivation

Technical objectives for sample surveys, of which a forest inventory is an example, include construction of inferences in the form of confidence intervals for population parameters. The Oxford English Dictionary defines the term *infer* as “to accept from evidence or premises” (Simpson & Weiner, 1989). For most scientific problems, evidence in the form of complete enumerations of populations of interest would be prohibitively expensive, if not physically impossible. Thus, statistical procedures have been developed to infer values for population parameters from estimates based on observations from a sample of population units. In this context, inference requires expression of the relationship between the population parameter, μ, and its estimate, $\hat{\mu}$, in probabilistic terms (Dawid, 1983). For situations in which the intent is estimation, as opposed to hypothesis testing, these probabilistic expressions often take the form of $1–\alpha$ confidence intervals,

$$\hat{\mu} \pm t_{1-\alpha} \cdot \sqrt{\hat{\mathrm{Var}}(\hat{\mu})}, \quad (1)$$

where $1–\alpha$ denotes the probability that confidence intervals constructed using data for all possible samples will include μ. Thus, the inference problem focuses on $\hat{\mu}$ and $\mathrm{SE}(\hat{\mu}) = \sqrt{\hat{\mathrm{Var}}(\hat{\mu})}$.

Two approaches to inference are relevant, the familiar probability- or design-based inference and model-based inference. Probability-based inference requires a probability sample and for sufficiently large samples produces estimates with acceptable precision. However, when only small samples can be acquired, particularly for highly variable populations, probability-based inference may fail to produce acceptably precise results. In addition, when no ground sampling is possible because the area of interest is remote or inaccessible and other information such as fine resolution remotely sensed data is lacking, probability-based inference is not possible. Examples include some tropical forests to be surveyed under the auspices of programs such as the United Nations initiative on Reducing Emissions due to Deforestation and Forest Degradation in developing countries and large, remote boreal regions such as interior Alaska in the United States of America (USA).

A general consensus is that inference for remote and inaccessible regions must rely on remotely sensed data, possibly in the form of maps. Of importance, however, maps only rarely accurately depict populations and provide no direct estimates of population parameters that are the primary survey objectives. Further, even if map unit predictions are aggregated to produce an estimate, map accuracy indices provide no direct information regarding the bias of the estimator resulting from classification and prediction errors or the precision of the estimate (McRoberts, 2011) and, therefore, cannot directly contribute to constructing inferences.

http://dx.doi.org/10.1016/j.rse.2014.08.028
0034-4257/Published by Elsevier Inc.

An alternative form of inference, characterized as model-based inference, has the potential to circumvent at least some of the difficulties associated with survey inference for remote and inaccessible regions. The validity of model-based inference is based on correct model specification rather than probability samples. Therefore, when combined with synthetic estimation which uses information external to the area of interest (Särndal et al., 1992), model-based inference can be used for remote and inaccessible regions for which probability samples are logistically difficult or financially prohibitive.

1.2. Objectives

The primary objective was to compare estimates obtained using model-based inference for a study area lacking sample data to estimates obtained using both model-based and model-assisted inference for the same study area when sample data were available. For a study area in Minnesota, USA, Landsat data were used to construct inferences for proportion forest area, and for a study area in Hedmark County, Norway, airborne laser scanning (ALS) data were used to estimate mean above-ground biomass per unit area (AGB).

2. Data

2.1. Minnesota, USA, study area

The study area was defined by the portion of the row 27, path 27, Landsat scene in northern Minnesota, USA, that was cloud-free for 16 July 2002 (Fig. 1). The 30-m × 30-m image pixels served as population units. Four smaller, 700-km^2 areas of interest (AOI) within the study area were also selected. Spectral data in the form of the normalized difference vegetation index (NDVI) transformation (Rouse, Haas, Schell, & Deering, 1973) and the three tasseled cap (TC) transformations (brightness, greenness, and wetness) (Crist & Cicone, 1984; Kauth & Thomas, 1976) were used as auxiliary information.

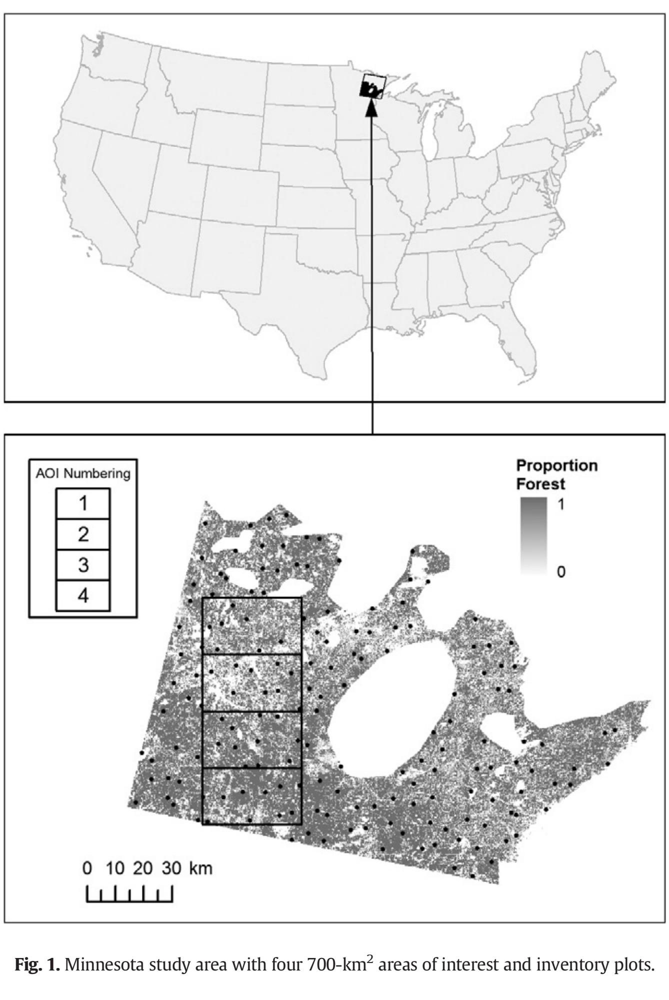

Fig. 1. Minnesota study area with four 700-km^2 areas of interest and inventory plots.

Ground data were obtained for plots established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service which conducts the national forest inventory (NFI) of the USA. The FIA program has established field plot centers in permanent locations using a sampling design that is regarded as producing an equal probability sample. Each FIA plot consists of four 7.32-m (24-ft) radius circular subplots that are configured as a central subplot and three peripheral subplots with centers located at distances of 36.58 m (120 ft) and azimuths of 0°, 120°, and 240° from the center of the central subplot (McRoberts, Bechtold, Patterson, Scott, & Reams, 2005; McRoberts, Hansen, & Smith, 2010). In general, centers of forested, partially forested, or previously forested plots are determined using global positioning system (GPS) receivers with accuracies of 10 m or greater, and centers of non-forested plots are verified using aerial imagery and digitization methods. Field crews visually estimate the proportion of each subplot that satisfies the FIA definition of forest land: minimum area of 0.4 ha (1.0 ac), minimum crown cover of 10%, minimum crown cover width of 36.6 m (120 ft), and forest land use. Subplot-level proportion forest was combined with the values of the spectral transformations for pixels containing subplot centers.

Because the smaller 168.3-m^2 subplots may not adequately characterize the larger 900-m^2 TM pixels, subplots whose observations were not completely forested or completely non-forested were deleted and assumed to be missing at random. In addition, to avoid issues related to spatial correlation among observations of subplots of the same plot, data for only the central subplot of each plot were used for this study. Subsequent to deletions, data for 168 plots measured in 2002 were available for the study. For future reference, the term *plot* refers to the central subplot of each FIA plot cluster.

2.2. Hedmark, Norway, study area

The study area was in the municipalities of Åmot and Stor-Elvdal in Hedmark County, Norway (Fig. 2). Four smaller, 100-km^2 AOIs within the study area were selected. The entire study area was tessellated into square 250-m^2 cells that served as population units.

ALS data were acquired between 15 July 2006 and 12 September 2006 with average density of 0.7 pulses/m^2. For each plot and population unit, height distributions were estimated for first echoes with heights greater than 2 m, and two sets of ALS metrics were calculated (Gobakken & Næsset, 2008). The first set of metrics consisted of heights corresponding to the 10th, 20th, …,100th percentiles of the distributions which were denoted h_{10}, h_{20}, …, h_{100}, respectively. The second set of metrics consisted of canopy densities calculated as the proportions of the same echoes with heights greater than 0%, 10%, …, 90% of the range between 2 m above the ground and the 95th percentile height and were denoted d_0, d_{10}, …,d_{90}, respectively.

Field measurements were obtained for 145 circular 250-m^2 Norwegian NFI field plots measured between 2005 and 2007. On each plot, all trees with diameters at-breast-height (dbh, 1.3 m) of at least 5 cm were callipered. Heights were measured on an average of 10 sample trees per plot selected with probability proportional to stem basal area, and heights were predicted using height-dbh models for trees whose heights were not measured. AGB was estimated at the plot-level using models, and any model prediction errors were ignored. Differential Global Navigation Satellite Systems (GPS and the Russian GLONASS) were used to determine the positions of the centers of plots with accuracies on the order of a few cm. Gobakken et al. (2012) describe this dataset in greater detail.