Forest Ecology and Management 401 (2017) 295–303

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Multivariate inference for forest inventories using auxiliary airborne laser scanning data

Ronald E. McRoberts [a,*], Qi Chen [b], Brian F. Walters [a]

[a] Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA
[b] Department of Geography, University of Hawai'i at Mānoa, Honolulu, HI, USA

ARTICLE INFO

Article history:
Received 12 April 2017
Received in revised form 8 July 2017
Accepted 10 July 2017
Available online 22 July 2017

Keywords:
k-Nearest Neighbors
Stratified estimator
Model-assisted regression estimator

ABSTRACT

National forest inventories have a long history of using remotely sensed auxiliary information to enhance estimation of forest parameters. For this purpose, aerial photography and satellite spectral data have been shown to be effective as sources of information in support of stratified estimators. These spectral-based stratifications are much more effective for reducing variances for forest area-related parameters than for parameters related to continuous attributes such as volume and biomass. For variables related to the latter attributes, stratified estimators using airborne laser scanning auxiliary data are much more effective, but are less effective than model-assisted estimators using the same auxiliary data. For inventory applications, however, stratified estimators using the same stratification for all response variables are naturally multivariate, whereas model-assisted estimators are not. A consequence is that multiple, univariate applications of model-assisted estimators cannot ensure compatibility among estimates of inventory parameters related to variables such as forest area, growing stock volume, and tree density.

The objectives of the study were twofold: (1) to optimize a multivariate, k-NN approach for simultaneously predicting multiple forest inventory variables; and (2) to compare multivariate model-assisted generalized regression estimators using optimized k-NN predictions to post-stratified estimators with respect to inferences in the form of confidence intervals for multiple forest inventory parameters. The analyses included use of airborne laser scanning data as auxiliary information and the multivariate k-NN technique for prediction in support of the model-assisted estimators. The study area was in north central Minnesota in the USA and is characterized by both lowland and upland forest types interspersed with wetlands and lakes.

The first primary result was that the optimized k-NN technique in combination with a model-assisted estimator produced compatible multivariate estimates of population means for six inventory parameters. Second, variances for the multivariate model-assisted estimators were smaller by 23%–35% than variances for a post-stratified estimator. These results warrant serious consideration of this approach for operational implementation by national forest inventories.

Published by Elsevier B.V.

1. Introduction

National forest inventories (NFI) in the Nordic countries and the United States of America (USA) have a long history of using remotely sensed auxiliary information to enhance inferences in the form of confidence intervals for forest inventory parameters. Aerial photography served as the earliest source of remotely sensed information for constructing stratifications for this purpose. Bickford (1953, 1960) in the USA and Poso (1972) in Finland constructed strata based on interpreted aerial photography to support stratified estimation. More recently, satellite imagery has served as the source of information for constructing stratifications. With this approach, the imagery is classified with respect to a forest attribute of interest, and the classes, or aggregations of the classes, serve as strata (Poso et al., 1984, 1987; Hansen and Wendt, 2000; McRoberts et al., 2002, 2006). McRoberts et al. (2012) provide more details on the history of using aerial photography and satellite imagery to support stratified estimators for forest inventory applications. Although stratifications based on aerial photography and satellite imagery have been shown to be effective for increasing the precision of estimators of forest area, their effectiveness for attributes such as growing stock volume and biomass is considerably less (McRoberts et al., 2006).

* Corresponding author.
E-mail address: rmcroberts@fs.fed.us (R.E. McRoberts).

http://dx.doi.org/10.1016/j.foreco.2017.07.017
0378-1127/Published by Elsevier B.V.

The advent of airborne laser scanning (ALS) data has introduced new possibilities for using remotely sensed auxiliary information to increase the precision of estimators of parameters related to forest volume and biomass. Næsset (2002) reported that 80–93% of the variability in field measured forest volume could be explained by models that use ALS metrics, and Næsset and Gobakken (2008) reported that 88% of the variability in aboveground biomass could be explained with models using similar metrics. These results have been confirmed in multiple additional studies (Li et al., 2008; Zhao et al., 2009; Frazer et al., 2011). McRoberts et al. (2012, 2013) demonstrated that ALS-based stratifications increase precision for estimators of growing stock volume comparable to the increases satellite image-based stratifications produce for forest area. However, for continuous forest attributes such as growing stock volume, model-assisted estimators using ALS data increase precision by even more than stratified estimators (McRoberts et al., 2013). With model-assisted estimators, an initial estimate based on model predictions for all population units is adjusted using differences between sample unit observations and predictions to compensate for systematic prediction error.

For operational purposes, NFIs require compatibility among estimates of parameters for different attributes. For example, for a particular estimation unit, a small estimate of forest area should not accompany a large estimate of growing stock volume. Such problems do not arise with stratified estimators using the same stratification because the stratifications only provide weights which are applied equally to observations for all response variables. Model-assisted estimators, on the other hand, require predictions for all response variables for all population units, and if a multivariate prediction approach is not used, then inevitably incompatible predictions such as large growing stock volume for a population unit that is predicted to have no forest cover will occur.

Multivariate regression methods typically require multivariate normally distributed response variables, a condition that is seldom satisfied for forest inventory variables. An alternative that has become very popular for use with remotely sensed data for forest inventory applications is the multivariate, non-parametric k-Nearest Neighbors (k-NN) technique (Chirici et al., 2016). Among the reported multivariate applications of k-NN, Temesgen et al. (2003) and LeMay and Temesgen (2005) predicted basal area and tree density using variables that included crown closure, height, age and ecological zone. McRoberts et al. (2007) and McRoberts (2009) predicted basal area, tree density and volume using Landsat metrics and used model-based inference to estimate small area means and their standard errors. Nothdurft et al. (2009) and Breidenbach et al. (2010) predicted total and three species-specific timber volumes for stands using ALS and optical data. Dash et al. (2015) predicted basal area, tree density, volume, and height using lidar metrics and estimated stand-level means and standard errors using the same approach to model-based inference. These studies established the utility of k-NN for multivariate prediction and for small area, model-based inference. However, none of these studies focused on larger areas on the order of inventory reporting units that are amenable to probability-based (design-based) inferential methods.

The objectives of the study were twofold: (1) to optimize a multivariate, k-NN approach for simultaneously predicting multiple forest inventory variables; and (2) to compare multivariate model-assisted generalized regression (GREG) estimators using optimized k-NN predictions to post-stratified (STR) estimators with respect to inferences in the form of confidence intervals for multiple forest inventory parameters. For both the stratified and model-assisted estimators, the auxiliary information was in the form of metrics derived from ALS data. The study area was in north central Minnesota in the USA and is characterized by both lowland and upland forest areas interspersed with wetlands and lakes.

2. Data

2.1. Study area

The 7583-km^2 study area consisted of the entirety of Itasca County in north central Minnesota in the USA (Fig. 1). Land cover includes water, wetlands and forest consisting of uplands with deciduous mixtures of pines (*Pinus* spp.), spruce (*Picea* spp.), and balsam fir (*Abies balsamea* (L.) Mill.) and lowlands with spruce (*Picea* spp.), tamarack (*Larix laricina* (Du Roi) K. Koch), white cedar (*Thuja occidentalis* (L.)), and black ash (*Fraxinus nigra* Marsh.).

2.2. Forest inventory data

Data were obtained for plots established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service which conducts the NFI of the USA. The FIA program has established field plot centers in permanent locations using a systematic unaligned sampling design that is regarded as producing an equal probability sample (McRoberts et al., 2010). The entire array of plots for Minnesota is subdivided into five systematic interpenetrating panels, and one panel is selected on a rotating basis for measurement each year. 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 36.58 m (120 ft) and azimuths of 0°, 120°, and 240° from the center of the central subplot. Field crews visually estimate the proportion of each subplot that satisfies the FIA definition of forest land: (i) minimum area 0.4 ha (1.0 ac), (ii) minimum tree cover of 10%, (iii) minimum width of 36.58 m (120 ft), and (iv) forest land use. For plots on forest land, field crews also observe species and measure diameter at breast-height (dbh, 1.37 m, 4.5 ft) and height for all trees with dbh of at least 12.7 cm (5 in.) on each subplot. Allometric model predictions of individual tree stem volumes are aggregated at subplot-level. For this study, uncertainty associated with the allometric model predictions was ignored. Species-level specific gravities are used to convert tree volumes to aboveground live tree stem biomass. Subplot-level response variables for this study included proportion forest area (A), basal area (BA, m^2/ha), growing stock volume (V, m^3/ha), aboveground live tree stem biomass (AGB, Mg/ha), tree density (D, stems/ha), and mean live tree height (HT, m).

Data were used for only the central subplots of the 242 plots measured in 2014 and 2015, because these were the only subplots and years for which plot coordinates were obtained using survey grade GPS receivers with sub-meter accuracy. For further reference, use of the term *plot* refers to the central subplot.

2.3. Airborne laser scanning data

Wall-to-wall ALS data were acquired in April 2012 with a nominal pulse density of 0.67 pulses/m^2. Ground returns were classified by the provider and were used to construct a digital terrain model via interpolation using the Tiffs (Toolbox for Lidar Data Filtering and Forest Studies) software (Chen, 2007). For this study that uses relatively small plots and ALS data characterized by small pulse densities, all pulse returns were used.

Distributions of all pulse return heights were constructed for the 168.3-m^2 plots and for the 169-m^2 square cells that tessellated the study area and served as population units. ALS metrics for each plot and cell included the mean (h_{mn}), standard deviation (h_{sd}), skewness (h_{sk}), kurtosis (h_{ku}), and quadratic mean height (h_{qm}) of the distributions of heights for all pulse returns (Lefsky et al., 1999; Chen et al., 2012). In addition, heights corresponding to the 10th, 20th, ..., 100th percentiles (h_{10}, h_{20}, ..., h_{100}) of the distributions were calculated as were canopy densities expressed as