ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Hierarchical Bayesian models for small area estimation of forest variables using LiDAR

Neil R. Ver Planck[a,*], Andrew O. Finley[a], John A. Kershaw Jr.[b], Aaron R. Weiskittel[c], Megan C. Kress[d]

[a] Department of Forestry, Michigan State University, East Lansing, MI 48824-1222, USA
[b] Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, NB E3B 5A3, Canada
[c] School of Forest Resources, University of Maine, Orono, ME 04469, USA
[d] Departments of Computer Science and Forestry, Michigan State University, East Lansing, MI 48824-1222, USA

ARTICLE INFO

Keywords:
Aboveground forest biomass
Area-level
Hierarchical Bayes
LiDAR
small area estimation

ABSTRACT

Light detection and ranging (LiDAR) data have become almost ubiquitous as a remote sensing tool in forestry estimation and mapping applications. Such initiatives commonly rely on spatially aligned forest inventory plot measurements and LiDAR covariates to inform model-based estimators for small area estimation. There are many examples where such linking models provide the desired accuracy and precision of forest parameter estimates for small areas where paucity of inventory plot observations preclude design-based inference. This paper builds on previous small area estimation (SAE) work by linking LiDAR covariates with variable radius forest inventory plot measurements within a hierarchical Bayesian framework. Using this framework, we compare SAE of forest aboveground biomass using: *i*) Fay-Herriot (FH); *ii*) FH with conditional autoregressive random effects (FHCAR); and *iii*) FHCAR with smoothed sampling variance (FHCAR-SMOOTH) models. Candidate models and the direct estimate based on plot measurements alone were compared using coefficient of variation (CV). On average, the FH model reduced the CV by 52.3% compared to the direct estimate. Incorporating spatial structure via the FHCAR model reduced the CV by 56.9% and 10.8% relative to the direct and the FH model estimates, respectively. Overall, these results illustrate the applicability and utility of using a SAE framework for linking LiDAR with typical forest inventory data.

1. Introduction

Forest inventory efforts typically follow a sampling design that aims to cover a potentially broad range of stand conditions, e.g., capturing species and structural diversity. Generally, a systematic grid of fixed-area plots (FAP) or variable-radius plots (VRP) are established across the forest. The choice between FAP or VRP depends on the inventory objectives (Maltamo et al., 2009). A greater cost efficiency generally results in the establishment of VRP for operational inventories over FAP research inventories (Rice et al., 2014).

Light detection and ranging (LiDAR) data have become one of the remote sensing tools of choice for extending and improving ground-based forest inventories and monitoring. There has been considerable research relating forest attributes with LiDAR covariates (see reviews by McRoberts et al., 2010; Næsset et al., 2004). In the case of complete LiDAR coverage, a fine grid is typically imposed on the area of interest and LiDAR covariates are calculated using the point cloud within each grid pixel. Often there is some effort to matching the pixel size to that of the inventory plot, especially in the setting where a regression model is developed to relate the LiDAR covariates to the response forest variables of interest (see, e.g., Finley et al., 2013; McRoberts et al., 2013). Development of regression models for spatially aligned LiDAR and inventory plot measurements is often referred to as *unit-level* analysis and is quite common (see, e.g., Babcock et al., 2015, 2016; Finley et al., 2017; Gregoire et al., 2016). These and similar analyses use FAP, opposed to VRP, because the relationship between plot extent can be directly matched with spatially coinciding LiDAR covariates. Truncated VRP, which result in a comparable extent to FAP, used in the Finnish National Forest Inventory have been successfully regressed on spatially aligned LiDAR covariates (Maltamo et al., 2007), although this may not generalize to stands with diverse diameters (Scrinzi et al., 2015).

There are examples, where VRP have been used in unit-level analyses (Hollaus et al., 2007, 2009). More recently, Hayashi et al. (2015) and Deo et al. (2016) have applied a variety of LiDAR resolutions that best match with the basal area factor applied in the VRP sampling. As an alternative to a unit-level analysis, an *area-level* analysis can be used

* Corresponding author.
E-mail address: verplan6@msu.edu (N.R. Ver Planck).

http://dx.doi.org/10.1016/j.rse.2017.10.024
Received 30 April 2017; Received in revised form 4 October 2017; Accepted 14 October 2017

when there is spatial misalignment between LiDAR and plot measurements (Goerndt et al., 2011). In an area-level regression analysis, observations are at the stand-level, whereas a unit-level analysis could consider multiple observations within each stand. Working at the area-level affords some advantages. For example, one can combine the cost efficient VRP with LiDAR covariates without spatial alignment between the plot data and LiDAR (Goerndt et al., 2011). In an effort to align VRP measurements and LiDAR data, Kronseder et al. (2012) calculated LiDAR covariates at a resolution of 1 ha based on the idea that VRP measurements are expanded to a per ha basis. Additionally, van Aardt et al. (2006) and Hudak et al. (2014) calculated the LiDAR variables at a segment- or stand-level in which the VRP samples were established. In this study, our focus is on application of area-level models to estimate aboveground forest biomass (AGB), with inference at the stand-level. The area-level models developed here are general and could be applied to other forest variables or transformations of AGB, e.g., for use in forest carbon accounting projects.

Given time and cost constraints, inventory designs often focus on achieving forest-level accuracy and precision requirements, which results in a limited number of samples collected within any given stand. These small sample sizes result in stand-level point estimate uncertainty that is too large for practical use. This limited inference at the stand-level, due to paucity of samples, is referred to as the small area problem and is commonly tackled using a small area estimation (SAE) method. We consider a forest stand to be the small area of interest for the following development; however, this is setting specific and one can of course consider other small area delineations.

The key of SAE is to borrow strength across related areas to improve estimation at the small area of interest. Rao and Molina (2015) give an in-depth overview of SAE and cover many different approaches. Following the definitions of Rao and Molina (2015), SAE is divided into three classes: *i*) direct; *ii*) indirect; and *iii*) small area model-based estimation. Direct estimators are generally design-unbiased for the area of interest and derived directly from the sample. Model-assisted methods, also belonging to this class, have proven useful for combining remotely sensed data with ground observations for AGB estimation (see, e.g., Opsomer et al., 2007). Indirect estimators are developed based upon an implicit model that borrows strength from either another domain, time, or both. Several frequently applied examples of this class are synthetic and composite estimators (Breidenbach and Astrup, 2012; Goerndt et al., 2013, 2011). In this study, we consider the third class of SAE, small area model-based estimation, referred to as small area models. Small area models differ from the previous two classes by including an explicit model with random effects that account for variation not explained by covariates in the model mean.

For small area models, the analyst applies either a frequentist or Bayesian mode of inference. The frequentist approach uses empirical best linear unbiased prediction (EBLUP), and the Bayesian approach uses either empirical Bayes (EB) or hierarchical Bayes (HB). For forestry applications, EBLUP has been applied most frequently (Breidenbach and Astrup, 2012; Goerndt et al., 2013, 2011; Magnussen et al., 2014; Mauro et al., 2016). Mauro et al. (2016) emphasized the correct specification of the estimator of mean squared error (MSE) of EBLUP for different levels of aggregation, from a pixel to an entire forest. The most common small area model is a linear mixed effects model, called the Fay-Herriot (FH, Fay and Herriot, 1979) model, which links a direct estimator to covariates via a linear model. Only one of the preceding SAE forestry studies examined spatial correlations among the area-level effects (see Appendix B in Magnussen et al., 2014). EB is considered the Bayesian paradigm equivalent to EBLUP. Alternatively, HB methods provide access to posterior distributions of the small area parameters (You and Zhou, 2011), and hence parameter inference that does not rely upon potentially unrealistic asymptotic assumptions (Pfefferman, 2013).

The primary objective of this study was to apply a HB framework to increase the precision of estimates for mean AGB at the stand-level by borrowing strength across all stands through the use of LiDAR covariates. Additionally, we apply a conditional autoregressive structure to the stand-level random effects to assess gains in precision of AGB. The remainder of the manuscript follows with: *i*) a description of the study area along with relevant data for the small area models; *ii*) a description and implementation of the small area models; and *iii*) the results and discussion of applying small area models for AGB. All source code and data are provided to facilitate reproducible research and application of the proposed methods.

2. Methods

2.1. Data

2.1.1. Study area

The area of interest for this study was the Noonan Research Forest (NRF) near Fredericton, New Brunswick, Canada (N 45° 59′ 12″, W 66° 25′ 15″). The NRF has been managed by the University of New Brunswick since 1985 and is approximately 1500 ha in size with a total of 271 stands. The subsequent analysis uses a subset of 226 stands each with a minimum of two VRP per stand. These stands ranged in size from 0.6 to 47 ha with an average size of 6.6 ha (Table 1; Fig. 1). The forest is composed of hardwood, mixed, and softwood stands with the major species being aspen (*Populus* spp.), balsam fir (*Abies balsamea* L. (Mill.)), birch (*Betula* spp.), eastern white pine (*Pinus strobus* L.), red maple (*Acer rubrum* L.), and spruce (*Picea* spp.), see Hayashi et al. (2015) for more details.

2.1.2. Variable radius plot data

In 2010, a 100 x 100 m grid was laid out across the NRF. At each grid intersection a VRP was established and trees greater than 6.0 cm diameter at breast height (DBH) were selected into the sample using a 2 M basal area factor angle gauge. Species, DBH, and height were recorded for each sample tree. Plot estimates of AGB Mg ha^{-1} were calculated using Jenkins et al. (2003) species-group equations. Stand-level estimates were obtained by averaging plot-level AGB Mg ha^{-1} estimates. Table 1 summarizes these stand-level estimates.

2.1.3. LiDAR data

The full waveform LiDAR data were collected on October 21 and 22, 2011 using a Riegl LMS Q680i laser scanner mounted on an airplane. The sensor had a pulse repetition frequency of 180 kHz with a laser wavelength of 1550 nm and a scan angle $< 28.54°$ from nadir. The forest was covered in overlapping strips to achieve at a minimum of six pulses per m^2, footprint of 0.35 m, and up to eight returns per pulse (Hayashi et al., 2015).

Stand-level LiDAR covariates were computed using the lascanopy function in the LAStools software suite (Isenburg, 2016). The NRF stand polygons, LAS files, and arguments to define the vertical extent of the

Table 1
Summary statistics for stand area, number of plots, mean aboveground biomass, sampling variances, and LiDAR covariates for the Noonan Forest ($m = 226$ stands) dataset.

	Min	Max	Mean	SD
Stand area (ha)	0.6	47.3	6.1	5.6
No. of plots	2	44	5.9	5.5
Mean AGB (Mg ha^{-1})	16.9	223.5	117.8	44.8
σ_i^2	0.158	7948	1698	1432
$\tilde{\sigma}_i^2$	36.5	804	411	226
$P25$ (m)	2.2	8.4	5.2	1.3
$P75$ (m)	5.1	20.7	11.7	2.8