Remote Sensing of Environment 204 (2018) 287–295

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
Available online 06 November 2017
0034-4257/ © 2017 Elsevier Inc. All rights reserved.

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
$\widetilde{\sigma}_i^2$	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

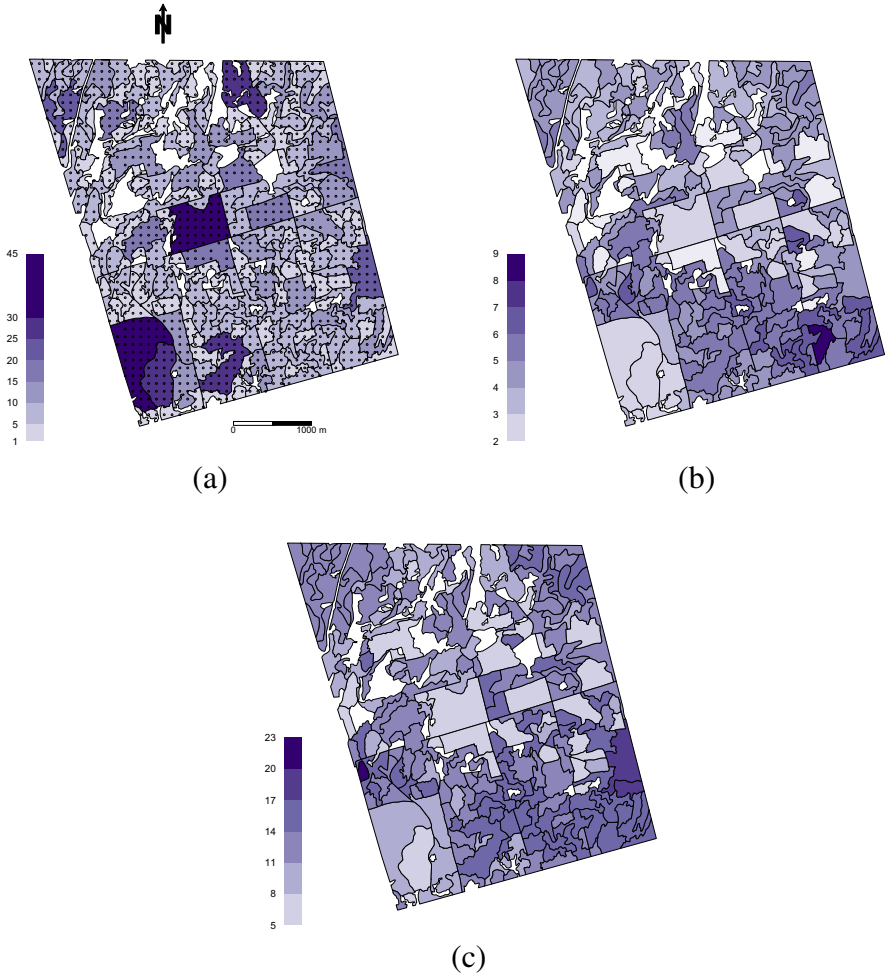

Fig. 1. Noonan Forest stands by (a) number of variable radius plots, (b) stand-level *P*25 (m); (c) stand-level *P*75 (m) LiDAR covariates.

LiDAR profile considered, were passed into the lascanopy function to generate the desired LiDAR covariates. Here, we consider stand-level LiDAR signal derived percentiles—height in meters at which some percent of the energy is returned—labeled *P* followed by the percentile, as well as signal shape variables, e.g., kurtosis and skewness. Exploratory analysis using a backward selection procedure resulted in *P*25 and *P*75 having the lowest Akaike information criterion value. Although the two percentile heights have a high correlation, the variance inflation factor of each was three, which is less than the general rule of ten for assessing multicollinearity (see Kutner et al., 2004, p. 409). Therefore, *P*25 and *P*75 covariates were used for the SAE models.

2.2. Models

2.2.1. Direct estimator and small area models

The direct estimator applied assumed simple random sampling for point and variance estimates of AGB. Beginning with the SAE FH model for stand i in 1,2,…,m stands defined as:

$$\begin{aligned} Y_i &= \theta_i + \epsilon_i, \\ \theta_i &= \mathbf{x}_i'\boldsymbol{\beta} + v_i, \end{aligned} \tag{1}$$

where Y_i is the direct estimate, the parameter of interest θ_i is mean AGB, and ϵ_i is a normally distributed error with mean zero and variance σ_i^2. The additive mean of θ_i comprises an intercept and stand-level covariates held in the $p \times 1$ column vector $\mathbf{x}$ and associated $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}$. The stand-level random effects term v_i is normally distributed with mean zero and variance σ_v^2. For comparison with HB inference, the frequentist EBLUP was also applied for the FH model with available R statistical software (see Appendix A).

It is reasonable to think that environmental variables, e.g., disturbance history and soil characteristics, could exhibit spatial autocorrelation and effect the observed AGB. In our setting, if direct estimate values are spatially correlated, i.e., adjacent stands have similar AGB values, then we should exploit this relationship to further improve inference by pooling information across proximate stands. Hence, we augment model Eq. (1) by adding a spatially structured random effect that follows a conditional autoregressive (CAR) prior distribution, see, e.g., Banerjee et al. (2015) and You and Zhou (2011). This extended model called FHCAR is defined analogous to FH, with the exception that the unstructured random effects $\mathbf{v} = (v_1, v_2, ..., v_m)' \sim N(\mathbf{0}, \sigma_v^2\boldsymbol{I})$ in model Eq. (1) are replaced with $\mathbf{v} \sim N(\mathbf{0}, \boldsymbol{\Sigma}(\sigma_v^2, \lambda))$. Here, the $m \times m$ covariance matrix $\boldsymbol{\Sigma}(\sigma_v^2, \lambda) = \sigma_v^2[\lambda\boldsymbol{R} + (1-\lambda)\boldsymbol{I}]^{-1}$, where σ_v^2 is the spatial variance parameter, λ is the autocorrelation parameter, $\boldsymbol{R}$ is the neighborhood matrix with diagonal elements equal to the number of neighbors and off diagonal elements equal negative one or zero indicating if a neighbor is present or not, and $\boldsymbol{I}$ is the $m \times m$ identity matrix. Stands are only considered neighbors with adjoining borders.

The previous models assume the sampling variances are fixed and known, which is a common assumption for SAE. However, additional smoothing methods are also commonly applied to the sampling variances to reduce instability in variance estimates of small sample sizes. Here, the sampling variances were smoothed by stand area a_i (similar to Goerndt et al., 2011); smoothed sampling variances $\tilde{\sigma}_i^2$ were defined as:

$$\tilde{\sigma}_i^2 = \frac{V_e}{n_i}; \quad V_e = \frac{\sum_{i=1}^{m} a_i \sigma_i^2}{\sum_{i=1}^{m} a_i}, \tag{2}$$

where n_i is the number of VRP in stand i. These $\tilde{\sigma}_i^2$ were then used in place of the original sampling variances in the FHCAR model. This model is referred to as the FHCAR-SMOOTH model. The development of generalized variance functions (GVF) are also common in the SAE literature (e.g., Dick, 1995); however, we did not find a useful covariate to develop a GVF.

The Bayesian SAE model specifications are completed by assigning prior distributions to parameters (Gelman et al., 2014). Each regression coefficient in $\boldsymbol{\beta}$ was assigned a flat prior distribution, σ_v^2 was given an inverse-Gamma (IG) prior distribution, and, following You and Zhou (2011), λ's prior was uniform with support between zero and one. The IG's shape hyperparameter was set to two, which results in a prior mean equal to the scale hyperparameter and infinite variance. The IG's scale hyperparameter was set as $\sum_{i=1}^{m} \sigma_i^2/m$ to give equal prior weight to the sampling and CAR variances. A Markov chain Monte Carlo (MCMC) algorithm was used to sample from parameters' posterior distributions. Specifically, a Gibbs algorithm was developed to sample from $\boldsymbol{\theta} = (\theta_1,\theta_2,\ldots,\theta_m)$, $\boldsymbol{\beta}$, and σ_v^2 with full conditional distributions given in You and Zhou (2011), and a Metropolis-Hastings algorithm was used to sample from λ's posterior distribution. We completed all analyses using `R` statistical software (R Core Team, 2014). The data and `R` code to reproduce this analysis are available at Mendeley Data (Ver Planck et al., 2017).

Parameter posterior inference was based on 3000 post burn-in MCMC samples from each of $L = 3$ chains. For the FH and FHCAR models, we thinned the samples to every third sample to reduce the autocorrelation in samples resulting in $K = 3000$ samples. For the FHCAR-SMOOTH model, thinning of the samples was not necessary based upon examination of plots of the autocorrelation function resulting in 9000 samples. Chain mixing and convergence were diagnosed using a multivariate potential scale reduction factor of less than 1.1 for all parameters considered (Gelman et al., 2014).

3. Results

The FH, FHCAR, and FHCAR-SMOOTH model parameter estimates are given in Table 2. Here, the 95% credible intervals for β_1 and β_2 do no include zero, which suggests the associated LiDAR covariates explain a substantial portion of the direct estimate's variability. The HB and EBLUP FH model parameter estimates were similar, with the exception of σ_v^2 (Table A1). The discrepancy between the HB and EBLUP estimate for σ_v^2 is likely due to the fact that, unlike EBLUP, HB acknowledges uncertainty in this parameter and hence greater variation around the AGB estimates.

The stand-level point estimates of AGB for the FH model ranged between 18.0 and 208.7 Mg ha^{-1} with a mean of 117.1 Mg ha^{-1}. The ranges of the FHCAR and FHCAR-SMOOTH models were 18.2–210.1 Mg ha^{-1} and 21.3–215.0 Mg ha^{-1} with means of 116.2 and 118.4 Mg ha^{-1}, respectively. Fig. 2 maps the direct and posterior means of AGB for FH, FHCAR, and FHCAR-SMOOTH model estimates for the individual stands. A scatter plot of these direct and posterior SAE model estimates does not reveal any regions of deviation (Fig. 3). Following the methods of Brown et al. (2001) and You and Zhou (2011), bias due to model misspecification was assessed through fitting linear models between the direct, assumed to be design-unbiased, and the posterior means of the SAE model estimates. The slopes of the individual linear models were 1.003, 1.002, and 1.063 for FH, FHCAR, and FHCAR-SMOOTH posterior means, respectively, which suggests there is not a systematic bias across the range of AGB values. The R^2 for each was 0.843, 0.838, and 0.890 for the FH, FHCAR, and FHCAR-SMOOTH posterior means.

The coefficient of variation (CV), defined as the square root of the posterior variance divided by the posterior mean, was used as the measure of precision among the direct, FH, FHCAR, and FHCAR-SMOOTH model estimates with a smaller value indicating better precision. For the direct estimates CV ranged between 0.004 and 1.156 with a mean of 0.363. The ranges for the FH, FHCAR, and FHCAR-SMOOTH models were 0.004–0.510, 0.004–0.398, 0.037–0.363 with means of 0.148, 0.133, and 0.106, respectively. Using a common CV threshold of 0.15 for stand-level estimates (Mauro et al., 2016), 29 of the 226 stands were at or below this threshold for the direct and 161 for the FH model estimates. The FHCAR and FHCAR-SMOOTH models had 168 and 195 stands, respectively, at or below the threshold. Fig. 4 maps the CV of the individual stands with the lightest shaded stands meeting or exceeding the CV threshold. Fig. 5 shows the CV of the direct and three model estimates against the stands ordered from smallest to largest CV of the direct estimator. Reductions in CV are observed between the direct estimator and the FH model for all but two stands, and even greater reductions are observed between the direct estimator and the FHCAR model. Further reductions in CV are observed between the FH and FHCAR model for 186 stands. The smoothing of the sampling variance for the FHCAR-SMOOTH model shows larger estimates of CV than even the direct estimator at the low end of the CV direct estimator. A total of 17 stands had a CV for the FHCAR-SMOOTH model greater than the direct estimator. These stands were generally stands with four or fewer variable radius plots that had low variation among the plots. The remaining stands' CV values showed a similar pattern to the FHCAR model.

The percent reductions in CV from the direct estimator to the FH model ranged from −0.9% to 88.0% with a mean of 52.3%. A wider range in percent reduction of −0.7%–90.7 % was observed from the direct estimator to the FHCAR model and a mean of 56.9%. Gains were also observed in percent CV reduction from the FH to the FHCAR model with a range of −34.6%–36.0 % and mean of 10.8%. The largest range in percent reduction was seen for the FHCAR to the FHCAR-SMOOTH model of −2220%–56.9 % and a mean of −1.2%. The average percent change across stands between the FHCAR and FHCAR-SMOOTH models was negative mainly due to large percent changes (< −1000%) in a few of the stands as depicted in Fig. 6. Excluding these two stands, the mean percent change in CV between the FHCAR and FHCAR-SMOOTH models was 13.9%.

Table 3 shows the mean percent reduction in CV for FH to FHCAR models and FHCAR to FHCAR-SMOOTH models by number of neighboring stands. For the FH to FHCAR models, a general increasing trend was observed in percent CV reduction with increasing number of neighbors (ranging from −25.2% to 36.0%) with a slight drop moving from ten to eleven neighbors. A similar pattern was observed for the FHCAR to FHCAR-SMOOTH models. The smaller number of neighbors (i.e., 1–3 and 5) had a negative value indicating reduction from FHCAR-SMOOTH to FHCAR (ranging from −35.1% to −21.8%), and the remaining number of neighbors showed reduction from FHCAR to

Table 2
The median and 95% credible intervals for Fay-Herriot (FH), FH with conditional autoregressive (FHCAR), and FHCAR with smoothed sampling variances (FHCAR-SMOOTH) model parameters.

Parameter	FH	FHCAR	FHCAR-SMOOTH
β_0	−51.99 (−69.15, −34.31)	−49.94 (−67.89, −31.02)	−41.50 (−54.77, −27.33)
β_1, $P25$	24.24 (18.17, 30.08)	23.35 (16.90, 29.52)	21.48 (16.80, 25.95)
β_2, $P75$	3.66 (0.83, 6.42)	3.82 (1.01, 6.84)	4.08 (2.28, 6.01)
σ_v^2	284.4 (201.2, 404.8)	471.5 (313.3, 720.8)	365.7 (201.3, 590.0)
λ	—	0.61 (0.21, 0.93)	0.61 (0.15, 0.93)

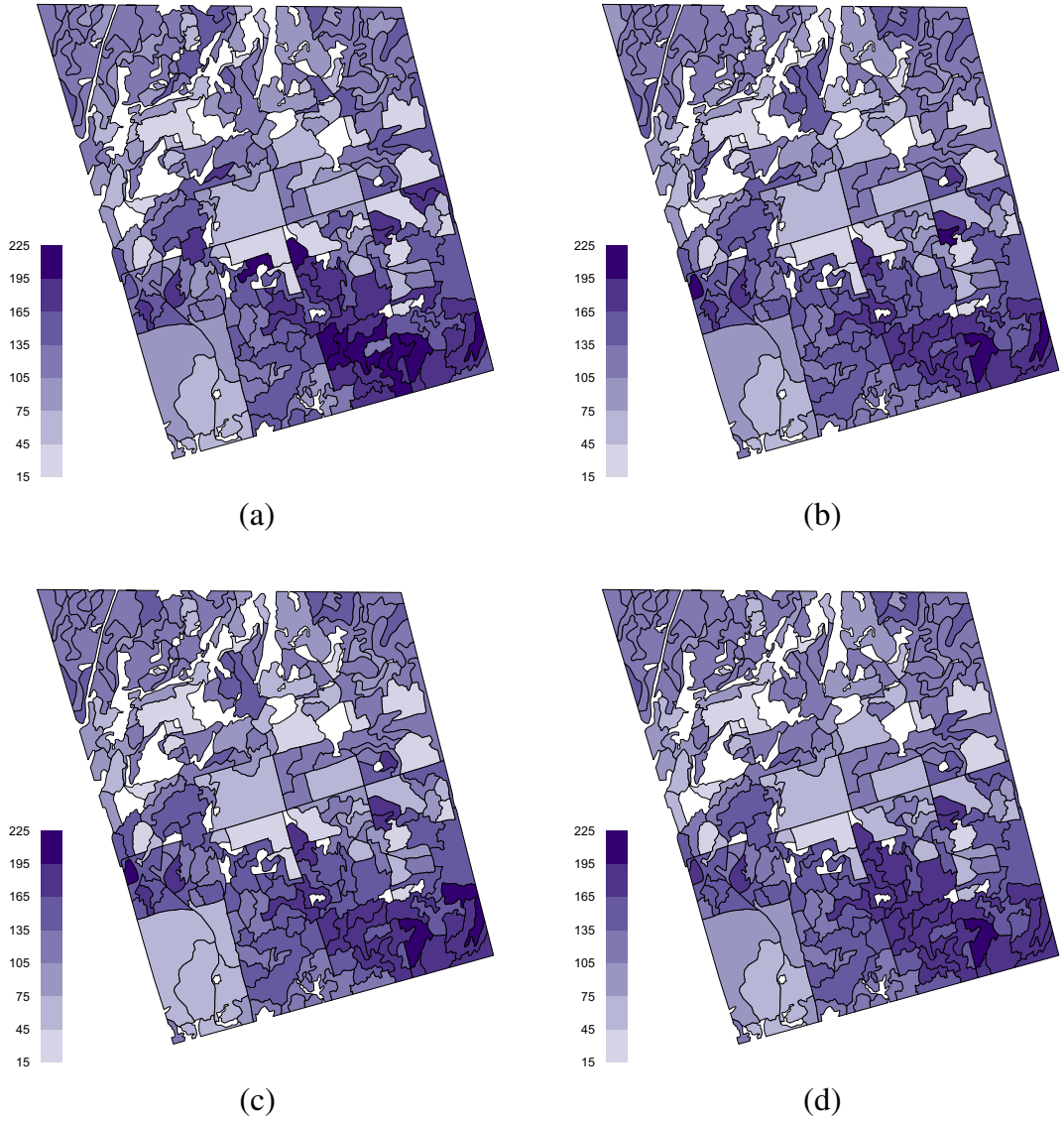

Fig. 2. Posterior mean aboveground biomass (Mg ha $^{-1}$) for the (a) direct estimate, (b) FH model, (c) FHCAR model; (d) FHCAR-SMOOTH model.

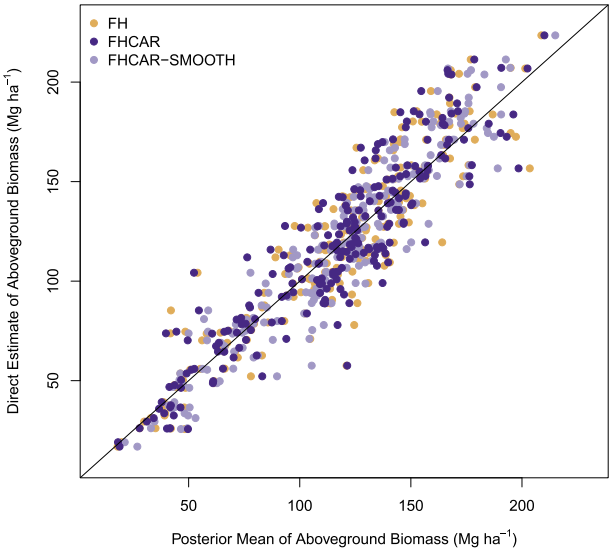

Fig. 3. Mean aboveground biomass (Mg ha $^{-1}$) for direct estimates versus posterior means of FH, FHCAR, and FHCAR-SMOOTH models.

FHCAR-SMOOTH (ranging from 20.3% to 43.9%).

The precision of the SAE models was also assessed by the width of the credible intervals, where narrower intervals indicate less uncertainty. The 95% credible interval widths of mean AGB are shown in Fig. 7 for the FH, FHCAR, and FHCAR-SMOOTH models. The credible interval widths for the FH model ranged from 1.6 to 70.8 Mg ha $^{-1}$ with a mean width of 57.2 Mg ha $^{-1}$. For the FHCAR model, the range of the credible interval widths were slightly wider, ranging from 1.5 to 87.3 Mg ha $^{-1}$, and the mean width was also smaller at 50.3 Mg ha $^{-1}$. The range for the FHCAR-SMOOTH model was smaller than the previous two models from 20.3 to 66.3 Mg ha $^{-1}$; however, the minimum credible interval width was much larger than both the FH and FHCAR models. The mean of the credible interval width of the FHCAR-SMOOTH model was 41.5 Mg ha $^{-1}$, which was slightly smaller than the FHCAR model.

4. Discussion

In this study, we found *P*25 and *P*75 to be the LiDAR covariates that explained the greatest variation in stand-level AGB. Hayashi et al. (2015) using the same data from NRF found the height of *P*45 to be most significant in explaining AGB estimates. One possible explanation

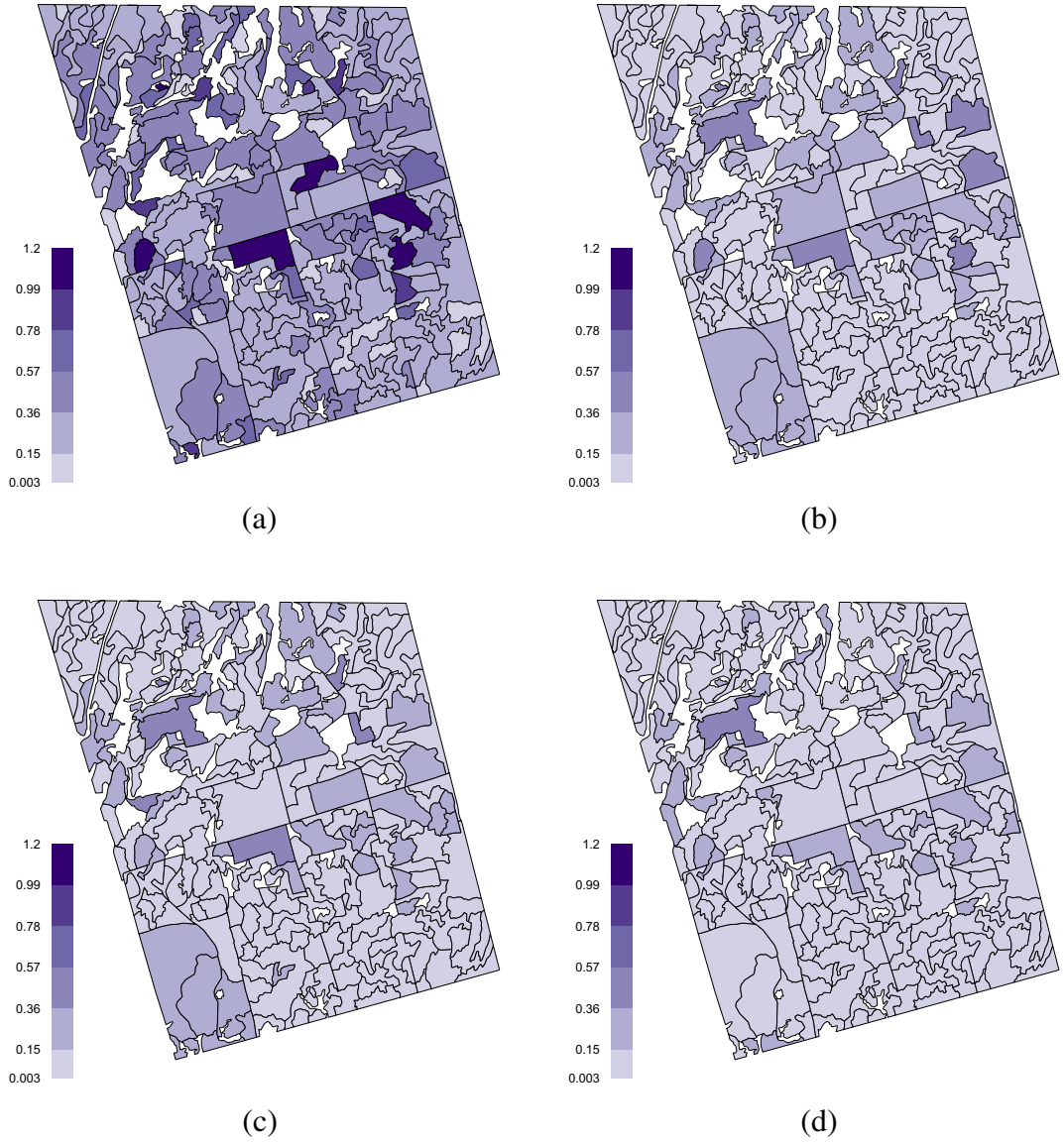

Fig. 4. Coefficient of variation of the (a) direct estimate, (b) FH model, (c) FHCAR model; (d) FHCAR-SMOOTH model.

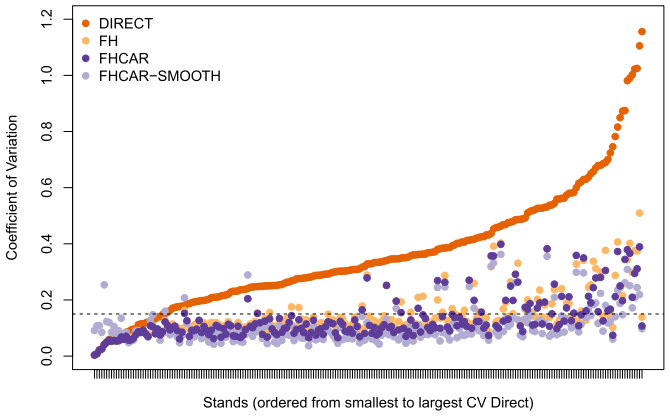

Fig. 5. Estimates of coefficient of variation (CV) versus stands ordered by increasing CV of direct estimate.

for this difference is the LiDAR covariates used in this study were summaries at the area-level (stand-level), while Hayashi et al. (2015) examined relationships between AGB and LiDAR covariates at the unit-level (plot-level). Goerndt et al. (2011) also found LiDAR covariates to differ in their final regression model for stand-level versus plot-level relationships. Many studies (e.g., Lefsky et al., 2002) have also found *P*95, generally regarded as canopy height, to be a common explanatory variable; however, this was not the case in our final linear model. We could have explored more LiDAR covariates; however, the focus of this paper was on applying the HB SAE framework and not to finding the best linking model.

Recently, Breidenbach et al. (2016) and Mauro et al. (2016) examined estimation of variance associated with model-based SAE from a frequentist perspective. In this study, we offered the hierarchical Bayesian method as an alternative that provided access to the full posterior distributions of the variables of interest and simplified interpretation of model parameter inference. For the FH model, the EBLUP had smaller σ_v^2 compared with HB. We see this as a strength of the HB

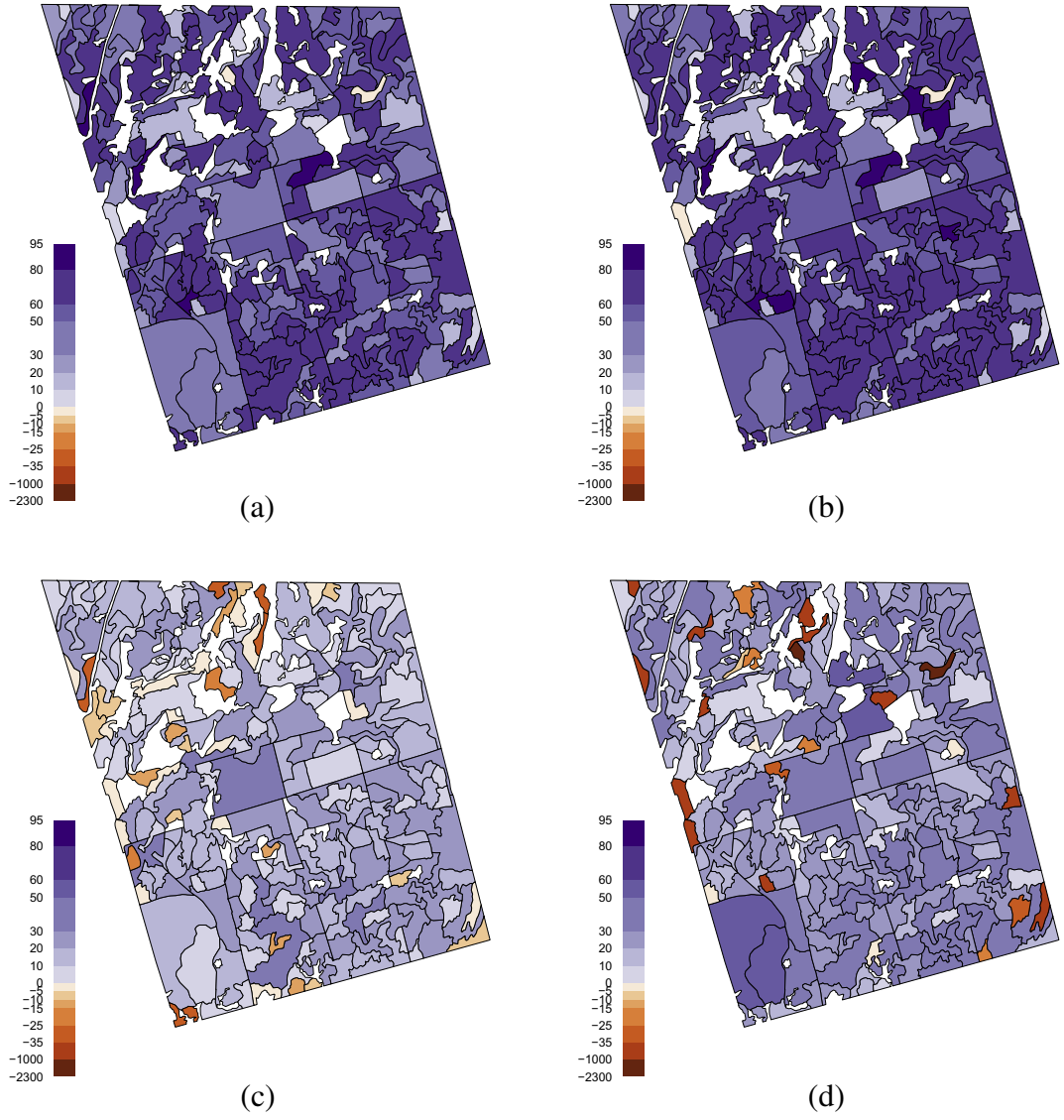

Fig. 6. Percent reduction in coefficient of variation from (a) direct estimate to FH model, (b) direct estimate to FHCAR model, (c) FH to FHCAR model; (d) FHCAR to FHCAR-SMOOTH model.

Table 3
Mean CV reduction (%) between SAE models by number of neighboring stands.

Number of neighbors	FH to FHCAR	FHCAR to FHCAR-SMOOTH
1	−25.2	−33.4
2	−7.3	−35.1
3	1.5	−21.8
4	9.3	21.6
5	15.3	−32.8
6	20.1	21.9
7	23.6	32.7
8	27.6	24.6
9	30.4	29.1
10	34.2	20.3
11	28.1	39.3
13	36.0	43.9

model, in that the HB mode of inference provides a more realistic accounting of uncertainty associated with estimating AGB. As a contribution, we have supplied the necessary functions to reproduce this research that could be applied to datasets with similar applications. This study also highlights the suitability of applying VRP for gains in sampling efficiency when coupled with LiDAR covariates that are summarized at the stand-level.

Large reductions in CV were seen in the application of either the FHCAR or FHCAR-SMOOTH model to the NRF stands due to strong area effects. Magnussen et al. (2014) also examined small area applications accounting for spatial autocorrelation with improvements for Swiss forest districts. Additionally, Breidenbach et al. (2016) found large improvements in coverage probability when incorporating spatial autocorrelation into the variance estimate for the confidence intervals. These examples show that, when present, accounting for spatial autocorrelation can better represent the uncertainty in the variables of interest.

Smoothing of sampling variances (Eq. (2)) in the FHCAR-SMOOTH model decreased credible interval widths compared to the FHCAR model; whereas, for stands with fewer neighboring stands and smaller sample sizes the FHCAR model had smaller credible interval widths. However, this may be a result of increasing the sampling variance for the smaller stands by applying smoothing and may be more reflective of the instability of variance with these small sample sizes. The method of

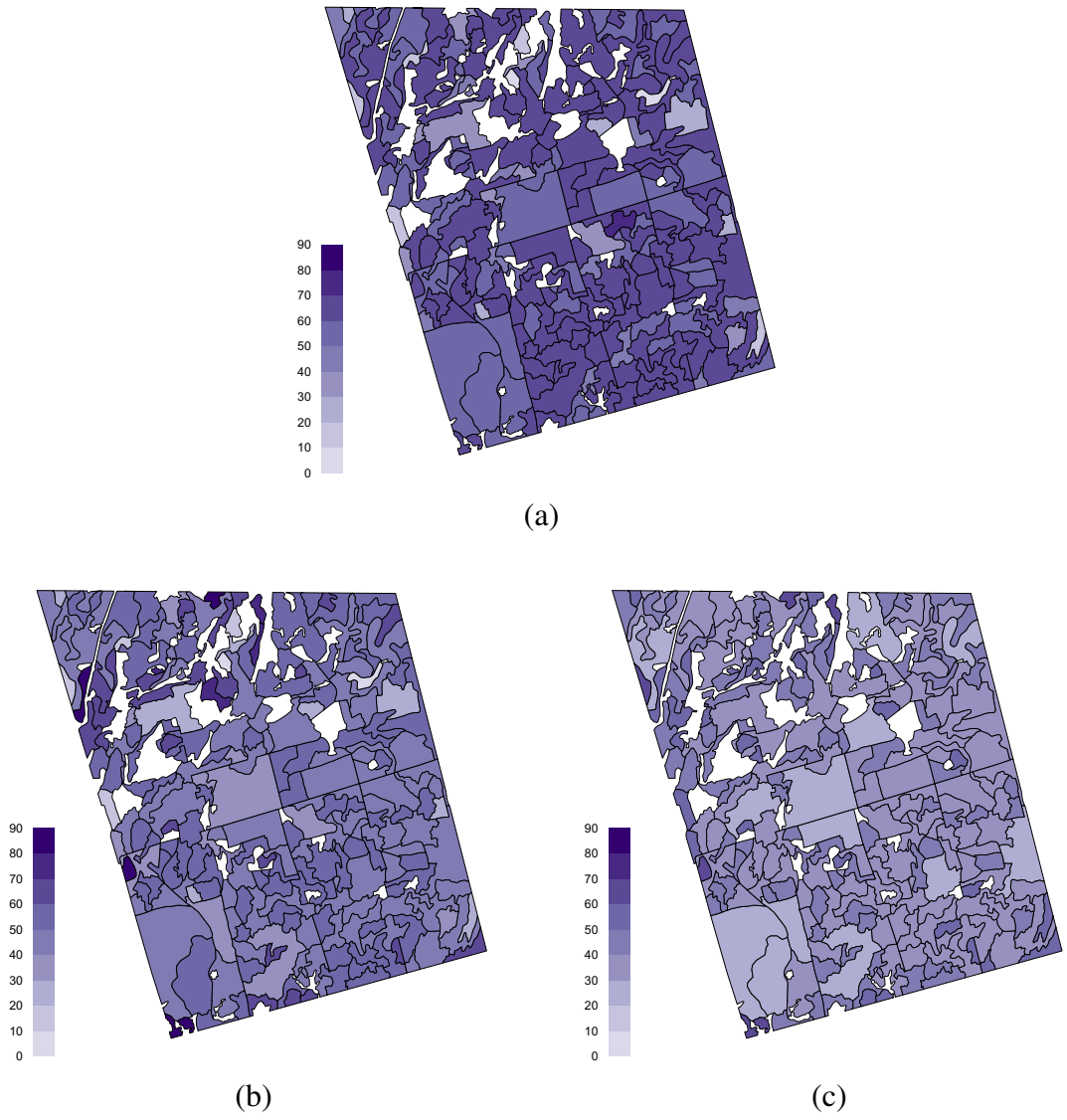

Fig. 7. 95% credible interval widths of AGB (Mg ha $^{-1}$) for the (a) FH, (b) FHCAR; (c) FHCAR-SMOOTH models.

smoothing based on stand area may not be reflective of increased variation in smaller stands due to the practice of delineating stands based on their similarities in management history, species composition, and age.

One potential drawback of applying SAE at the area-level is the inability to assess within stand variation, as can be done when aligning LiDAR and plot data at the unit-level (e.g., Hayashi et al., 2016; Woods et al., 2011). However, as seen in Hayashi et al. (2016) a determination of prediction unit size influences inference at the stand- and forest-level. For area-level analyses, the size of the areas can be adjusted to meet the application of interest with the requirement of at least two ground samples within each area for estimating the sampling variance.

A potential practical application of this research would be to extend AGB to carbon estimates for forest carbon accounting projects. SAE models could be cost effective in reducing uncertainty where paucity of field plots makes direct estimates unreliable for the project.

Additionally, future work will be to extend this study to the multivariate setting similar to the work of Porter et al. (2015). In this extension, cross-correlations in the multivariate setting could improve AGB estimates along with other common forest inventory variables, e.g., basal area, tree density, or volume. The HB SAE framework demonstrates the utility of applying common ground inventory techniques, i.e., sampling with VRP, with available remotely sensed data for reducing uncertainty at the area-level.

Acknowledgements

We would like to acknowledge the useful comments of two anonymous reviewers that improved the quality of this manuscript. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada, Discovery Grant programme (RGPIN 04280) and the New Brunswick Innovation Foundation, Research Assistants Initiative (NBIFRAI2013-073). Additional support came from NASA's Arctic-Boreal Vulnerability Experiment (ABoVE) and Carbon Monitoring System (CMS) grants. Andrew Finley was supported by National Science Foundation (NSF) DMS-1513481, EF-1137309, EF-1241874, and EF-1253225 grants.

Appendix A. HB and EBLUP parameter estimates for Fay-Herriot model

The empirical best linear unbiased prediction (EBLUP) for the Fay-Herriot model was fit via restricted maximum likelihood using the *hbsae* R package (Boonstra, 2012). The comparison in parameter estimates between the EBLUP and HB inference are given in Table A1.

Table A1

The median and 95% credible intervals for the model parameters of the Fay-Herriot (FH) with hierarchical Bayesian (HB) inference and the point estimates with standard errors in parentheses for the FH under empirical best linear unbiased predictor (EBLUP) fit via restricted maximum likelihood.

Parameter	FH HB	FH EBLUP
β_0	−51.99 (−69.15, −34.31)	−52.57 (8.52)
β_1, *P*25	24.24 (18.17, 30.08)	24.27 (2.95)
β_2, *P*75	3.66 (0.83, 6.42)	3.69 (1.39)
σ_v^2	284.4 (201.2, 404.8)	241.1

References

Babcock, C., Finley, A.O., Bradford, J.B., Kolka, R., Birdsey, R., Ryan, M.G., 2015. LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients. Remote Sens. Environ. 169, 113–127.

Babcock, C., Finley, A.O., Cook, B.D., Weiskittel, A., Woodall, C.W., 2016. Modeling forest biomass and growth: coupling long-term inventory and LiDAR data. Remote Sens. Environ. 182, 1–12.

Banerjee, S., Carlin, B., Gelfand, A., 2015. Hierarchical Modeling and Analysis for Spatial Data, second. CRC Press, Boca Raton, Florida 558 p.

Boonstra, H., 2012. hbsae: Hierarchical Bayesian Small Area Estimation. In: R Package Version 1.0, . http://cran.r-project.org/package=hbsae.

Breidenbach, J., McRoberts, R., Astrup, R., 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. 173, 274–281.

Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. 131, 1255–1267.

Brown, G., Chambers, R., Heady, P., Heasman, D., 2001. Evaluation of Small Area Estimation Methods—An Application to Unemployment Estimates from the UK LFS. Proceedings of Statistics Canada Symposium.

Deo, R.K., Froese, R.E., Falkowski, M.J., Hudak, A.T., 2016. Optimizing variable radius plot size and LiDAR resolution to model standing volume in conifer forests. Can. J. Remote. Sens. 42 (5), 428–442.

Dick, P., 1995. Modelling net undercoverage in the 1991 Canadian census. Surv. Methodol. 21, 45–54.

Fay, R.E., Herriot, R.A., 1979. Estimation of income for small places: an application of James-Stein procedures to census data. J. Am. Stat. Assoc. 74, 268–277.

Finley, A.O., Banerjee, S., Cook, B.D., Bradford, J.B., 2013. Hierarchical Bayesian spatial models for predicting multiple forest variables using waveform LiDAR, hyperspectral imagery, and large inventory datasets. Int. J. Appl. Earth Obs. 22, 147–160.

Finley, A.O., Banerjee, S., Zhou, Y., Cook, B.D., Babcock, C., 2017. Joint hierarchical models for sparsely sampled high-dimensional LiDAR and forest variables. Remote Sens. Environ. 190, 149–161.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian Data Analysis, third. Chapman & Hall/CRC Boca Raton, Florida 661 p.

Gregoire, T.G., Næsset, E., McRoberts, R.E., Ståhl, G., Andersen, H.-E., Gobakken, T., Ene, L., Nelson, R., 2016. Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. Remote Sens. Environ. 173, 98–108.

Goerndt, M.E., Monleon, V.J., Temesgen, H., 2013. Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. For. Sci. 59 (5), 536–548.

Goerndt, M.E., Monleon, V.J., Temesgen, H., 2011. A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. Can. J. For. Res. 41 (6), 1189–1201.

Hayashi, R., Weiskittel, A., Kershaw Jr., J.A., 2016. Influence of prediction cell size on LiDAR-derived area-based estimates of total volume in mixed-species and multi-cohort forests in northeastern North America. Can. J. Remote. Sens. 42 (5), 473–488.

Hayashi, R., Kershaw Jr., J.A., Weiskittel, A., 2015. Evaluation of alternative methods for using lidar to predict aboveground biomass in mixed species and structurally complex forests in northeastern North America. Math. Comput. For. Nat. Res. Sci. 7 (2), 49–65.

Hollaus, M., Wagner, W., Maier, B., Schadauer, K., 2007. Airborne laser scanning of forest stem volume in a mountainous environment. Sensors 7 (8), 1559–1577.

Hollaus, M., Wagner, W., Schadauer, K., Maier, B., Gabler, K., 2009. Growing stock estimation for alpine forests in Austria: a robust lidar-based approach. Can. J. For. Res. 39 (7), 1387–1400.

Hudak, A.T., Haren, A.T., Crookston, N.L., Liebermann, R.J., Ohmann, J.L., 2014. Imputing forest structure attributes from stand inventory and remotely sensed data in western Oregon, USA. For. Sci. 60 (2), 253–269.

Isenburg, M., 2016. LAStools — Efficient LiDAR Processing Software (Version 160721, Unlicensed). http://rapidlasso.com/LAStools.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49 (1), 12–35.

Kutner, M.H., Nachtsheim, C.J., Neter, J., 2004. Applied Linear Regression Models, fourth. McGraw-Hill/Irwin, New York, New York 701 p.

Kronseder, K., Ballhorn, U., Böhm, V., Siegert, F., 2012. Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. Int. J. Appl. Earth Obs. 18, 37–48.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. LiDAR remote sensing for ecosystem studies. Bioscience 52 (1), 19–30.

Magnussen, S., Mandallaz, D., Breidenbach, J., Lanz, A., Ginzler, C., 2014. National forest inventories in the service of small area estimation of stem volume. Can. J. For. Res. 44, 1079–1090.

Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L., Hyvönen, P., 2009. Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. Eur. J. For. Res. 128, 305–317.

Maltamo, M., Korhonen, K.T., Packalén, P., Mehtätalo, L., Suvanto, A., 2007. A test on the usability of truncated angle count sample plots as ground truth in airborne laser scanning based forest inventory. Forestry 80, 73–81.

Mauro, F., Molina, I., García-abril, A., Valbuena, R., Ayuga-Téllez, E., 2016. Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics 27 (4), 225–238.

McRoberts, R.E., Næsset, E., Gobakken, T., 2013. Inference for lidar-assisted estimation of forest growing stock volume. Remote Sens. Environ. 128, 268–275.

McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V., Tomppo, E.O., 2010. Using remotely sensed data to construct and assess forest attribute maps and related spatial products. Scand. J. Forest Res. 25 (4), 340–367.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., Söderman, U., 2004. Laser scanning of forest resources: the Nordic experience. Scand. J. Forest Res. 19 (6), 482–499.

Opsomer, J.D., Breidt, F.J., Moisen, G.G., Kauermann, G., 2007. Model-assisted estimation of forest resources with generalized additive models. J. Am. Stat. Assoc. 102 (478), 400–409.

Pfefferman, D., 2013. New important developments in small area estimation. Stat. Sci. 28 (1), 40–68.

Porter, A.T., Wikle, C.K., Holan, S.H., 2015. Small area estimation via multivariate Fay-Herriot models with latent spatial dependence. Aust. Nz. J. Stat. 57 (1), 15–29.

Core Team, R., 2014. R: a Language and Environment for Statistical Computing. In: R Foundation for Statistical Computing, Vienna, Austria.

Rao, J.N.K., Molina, I., 2015. Small Area Estimation, second. John Wiley & Sons, Inc., Hoboken, New Jersey 441 p.

Rice, B., Weiskittel, A.R., Wagner, R.G., 2014. Efficiency of alternative forest inventory methods in partially harvested stands. Eur. J. For. Res. 133 (2), 261–272.

Scrinzi, G., Clementel, F., Floris, A., 2015. Angle count sampling reliability as ground truth for area-based LiDAR applications in forest inventories. Can. J. For. Res. 45 (4), 506–514.

van Aardt, J.A.N., Wynne, R.H., Oderwald, R.G., 2006. Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segment basis. For. Sci. 52 (6), 636–649.

Ver Planck, N.R., Finley, A.O., Kershaw Jr, J.A., Weiskittel, A.R., Kress, M.C., 2017. [dataset] Hierarchical Bayesian Models for Small Area Estimation of Forest Variables Using LiDAR. Mendeley Data, v1. http://dx.doi.org/10.17632/vyfrnkx39h.1.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., Treitz, P., 2011. Operational implementation of a LiDAR inventory in boreal Ontario. Forest Chron. 87 (4), 512–528.

You, Y., Zhou, Q.M., 2011. Hierarchical Bayes small area estimation under a spatial model with application to health survey data. Surv. Methodol. 37 (1), 25–37.