LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients

Chad Babcock a, Andrew O. Finley b,⁎, John B. Bradford c, Randall Kolka d, Richard Birdsey e, Michael G. Ryan f

a School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA
b Departments of Forestry and Geography, Michigan State University, East Lansing, MI, USA
c US Geological Survey, Southwest Biological Science Center, Flagstaff, AZ, USA
d USDA Forest Service, Northern Research Station, Grand Rapids, MN, USA
e USDA Forest Service, Northern Research Station, Newtown Square, PA, USA
f Natural Resources Ecology Laboratory, Colorado State University, Fort Collins, CO, USA, and USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA

1. Introduction

Estimating forest above-ground biomass (AGB), along with other structure variables, using discrete Light Detection and Ranging (LiDAR) data is an active subject in ecological and resource monitoring research. Studies in this area show a strong potential for discrete return LiDAR to be used as a tool for developing spatially explicit estimates of many forest attributes, including AGB, either on its own or in conjunction with other remote sensing technologies (see, e.g., Gonzalez et al. (2010); Lim and Treitz (2004); Lucas et al. (2006); Sherrill, Lefsky, Bradford, and Ryan (2008)).

Regression models proposed for forest AGB mapping using LiDAR data often do not explicitly accommodate residual spatial dependence, see, e.g., Anderson et al. (2008), Gonzalez et al. (2010), Muss, Mladenoff, and Townsend (2011), Tonolli et al. (2011), and Popescu, Zhao, Neuenschwander, and Lin (2011). A non-spatial model can be appropriate if all spatially structured variation in the outcome is accounted for by the covariates used for model fitting; however, this is often an unrealistic assumption when data are spatially indexed. Considering forest attribute data, it is reasonable to expect similar outcomes for neighboring locations. For example, inventory plots close together could comprise like tree species and have comparable stem densities due to similar disturbance histories and environmental conditions. In contrast, inventory plots far apart are less likely to share common composition or structure attributes. A key assumption of regression is that model residuals are not correlated. Given the spatially dependent nature of forest inventory data, the potential for spatially correlated residuals in AGB models is high. Not accounting for residual spatial dependence can result in falsely precise estimates of model parameters and erroneous predictions (Hoeting, 2009). For these reasons, it is important to check for spatially structured residuals when spatial data are used to fit AGB models.
To ensure a model's statistical validity, spatial dependence among residuals can be accommodated by introducing a spatially varying intercept (SVI) via the addition of an appropriately structured random effect. Including spatial random effects that account for spatial association between observed locations can also improve model fit and prediction accuracy, via partitioning error uncertainty into a spatial and non-spatial component and borrowing information from observed locations to inform prediction at proximate locations.

In addition to a SVI, it may be appealing to allow the regression parameters associated with the covariates to vary across the study area. For instance, given the heterogeneity of forest species composition, age classes, and resources (e.g., light, water, soil characteristics), a single set of regression parameters might not adequately capture the space varying relationship between forest outcome variables and the covariates, i.e., parameters may have a spatially non-stationary relationship with the covariates. In such cases, one might attempt to capture these localized effects by including forest characteristics and/or environmental conditions as covariates in the regression model; however, it is not always clear which variables should be included and, in many applied settings, the necessary variables are not available. Rather, localized relationships between the outcome variable and covariates might be more effectively captured by allowing the regression coefficients to vary smoothly over the study area.

Such spatially varying coefficient (SVC) models have received some attention in the statistical literature (see, e.g., Finley, Banerjee, and McRoberts (2009); Finley, Banerjee, and MacFarlane (2011); Gelfand, Schmidt, Banerjee, and Sirmans (2004)), and more recently in applied disciplines (Finley, 2010; Wheeler & Calder, 2007; Wheeler & Waller, 2009). Flexible specification of these models follows the Bayesian paradigm of statistical inference (see, e.g., Banerjee, Carlin, and Gelfand (2004); Carlin and Louis (2000); Gelman, Carlin, Stern, and Rubin (2013)), where analysis uses samples, obtained via Markov chain Monte Carlo (MCMC) methods, from the posterior distributions of model parameters.

SVC and SVI models accommodate residual spatial dependence and parameter non-stationarity differently than more common geostatistical approaches seen in the remote sensing literature. Chen, Zhao, McDermid, and Hay (2012) examined geographically weighted regression (GWR, Fotheringham, Brunsdon, & Charlton, 2003) alongside other geostatistical approaches for predicting canopy height with remote sensing data and saw improvements in model fit. In an attempt to accommodate residual spatial dependence and parameter non-stationarity, Hudak, Lefsky, Cohen, and Berterretche (2002) explored kriging and co-kriging geostatistical methods to build prediction models for forest canopy height. They saw improvements in fit using regression kriging and regression co-kriging.

Recently, it has been shown that GWR may not be robust to correlation among covariates and produce erroneous results in the presence of complex correlation structures (Finley et al., 2011; Wheeler & Waller, 2009). Also, from an inferential standpoint, GWR can be problematic when attempting to draw inference about model parameters and prediction uncertainty. Because GWR lacks a valid underlying probability model, parameter standard error estimates and prediction variance estimates can potentially be unjustifiable. Error maps generated by kriging and co-kriging methods do not account for the uncertainty in the variogram-derived spatial covariance parameters, which is a well-known shortcoming encountered with these geostatistical approaches (Cressie, 1993).

In a Bayesian hierarchical framework, it is possible to estimate the spatial covariance parameters within SVI and SVC models and, hence, propagate their uncertainties through to the prediction of the outcome variable (Banerjee et al., 2004). Such approaches yield a more statistically defensible measure of uncertainty than would otherwise be produced using traditional kriging or GWR methods.

In an effort to more fully account for patterns of spatial dependence between AGB and LiDAR covariates, the utility of a Bayesian hierarchical modeling framework that accommodates both residual spatial dependence and non-stationarity of model covariates through the introduction of spatial random effects was assessed. This objective was explored using four forest inventory datasets that are part of the North American Carbon Program (NACP), each comprising point-referenced measures of AGB and discrete LiDAR. For each dataset, a set of regression model specifications of varying complexity was considered. Models were assessed based on fit criteria and predictive performance using a 10-fold cross-validation approach. AGB maps with associated pixel-level prediction uncertainty measures were generated using the model selected as “best” for each site. In cases where the selected model included spatial random effects, additional maps were created to illustrate features of the space-varying coefficients. Results showed, in an exploratory fashion, how uncertainty and space-varying coefficient maps can be used to identify: spatial extrapolation issues; missing covariates, and; locally significant regression coefficients.

The remainder of the paper evolves as follows. The motivating data and preprocessing steps are detailed in Section 2. The proposed modeling framework and model assessment are described in Section 3, followed by analysis results presented in Section 4. Finally, some discussion along with an indication of future work is provided in Section 5.

2. Study sites

2.1. Fraser Experimental Forest

The Fraser Experimental Forest (FEF) is located in central Colorado (39° 4’ N, 105° 52’ W) near the town of Fraser. Tree species at FEF consist primarily of Abies lasiocarpa Hook. Nutt and Picea engelmannii Parry at higher elevations and Pinus contorta Bol. at lower elevations. Climate at FEF is characterized by cold and relatively long winters, with mean annual temperature and precipitation of 0 °C and 737 mm, respectively.

FEF experienced a widespread stand-replacing fire in approximately 1685. Selective clearcuts were conducted at FEF as watershed-scale manipulations in the 1950s. Although FEF is currently experiencing mortality due to mountain pine beetle, the field measurements and LiDAR acquisition for this study were completed prior to beetle infestation.

2.2. Marcell Experimental Forest

Located in northern Minnesota, the Marcell Experimental Forest (MEF) consists of mixed forests that include both upland forests and peatlands. Upland forests are generally dominated by Pinus strobus Michx., but also contain substantial components of Betula papyrifera Marshall, Pinus resinosa Roezl, Pinus strobus L., and Pinus banksiana Lamb. Lowland tree species include Larix laricina (Du Roi) K. Koch, Picea mariana Britton, Sterns & Poggenb., Fraxinus nigra Marshall, and Thuja occidentalis L. Climate at MEF is subhumid continental, with mean annual precipitation of 785 mm, mean annual temperature of 3 °C and air temperature extremes from −46 °C to 38 °C. Forests of the Lake States region experienced widespread logging around the turn of the 20th century, including much of the MEF landscape. Natural disturbances at MEF include windstorms of variable intensity and rare wildfires.

2.3. Glacier Lake Ecosystem Experimental Site

Glacier Lake Ecosystem Experimental Site (GLEES) is located approximately 55 km west of Laramie, Wyoming at 3190 m elevation (41° 22’ N, 106° 14’ W). GLEES includes a mosaic of trees and alpine meadows; the forest is dominated by P. engelmannii and A. lasiocarpa. GLEES has a mean annual temperature of −2 °C and a mean annual precipitation of 1000 mm, primarily as snow. Tree ages at GLEES suggest either a stand-replacing disturbance more than 400 years ago with slow recovery, or a series of smaller disturbances over the last 400 years (Bradford, Birdsey, Joyce, & Ryan, 2008).
2.4. Niwot Long Term Ecological Research Site

Niwot Long Term Ecological Research Site (NIWOT) is located at 3050 m elevation on the front range of the Rocky Mountains (40° 2' N, 105° 33' W), near the town of Nederland, Colorado. Tree species include primarily a mix of *A. lasiocarpa*, *P. engelmannii* and *P. contorta* with minor components of *Pinus flexilis* A. Murr. and *P. tremuloides*. Mean annual temperature and precipitation are 4 °C and 800 mm, respectively. Disturbance history at NIWOT includes widespread clearcuts between 1900 and 1910.

2.5. Field data preparation

Field data at each site were collected using methods similar to the Forest Inventory and Analysis style plot design (Bechtold & Patterson, 2005). Each plot location consisted of 4 sub-plots with a radius of 8–10 m (depending on site) with 1 being in the center and 3 others 35 m away from the plot center at 0°, 120°, and 240°. Each site was initially established with 9 (FEF, GLEES, and NIWOT) or 16 (MEF) plots in 2004 located on a regular grid with 250 m between plot centers in the north-south and east-west directions. Additional plots and/or sub-plots were subsequently added during 2005 and 2006 at each site in locations selected to capture canopy height conditions not adequately represented in the initial gridded plots. FEF, MEF, GLEES and NIWOT contained 61, 115, 46 and 62 sub-plots respectively. Fig. 1 shows the spatial distribution of the sub-plots inside the LiDAR coverage area. Within each sub-plot, individual tree diameter at breast height, and height measurements for both live and dead trees were taken and used in species-specific allometric equations to estimate AGB (stem, branch and foliage biomass). Additional details about field measurements and allometric equations used for biomass estimation are available in Bradford et al. (2010) and Cole et al. (2013). The individual tree AGB estimates were totaled for each sub-plot and converted to AGB Mg/ha. The AGB measurements were then square root transformed to better approximate a normal distribution before model fitting.

2.6. LiDAR data preparation

Height-above-ground measurements for the first return pulses were calculated by subtracting the point elevations from a digital terrain model constructed from the LiDAR data. LiDAR return height empirical cumulative distribution curves were constructed over each sub-plot and a percentile height dataset was compiled by extracting heights associated with 5% intervals (i.e., 5%–100%) for each curve. These percentile heights served as an initial set of regression covariates. Using the same intervals, a percentile height dataset was constructed for a 20 × 20 m grid over the LiDAR coverage area and subsequently used to construct AGB prediction maps. The cell size was chosen because it is approximately the same area as the observed sub-plots.

Variable selection via dimension reduction was necessary because of high collinearity among the percentile height covariates. We chose principal components analysis (PCA) via eigen (spectral) decomposition of the percentile height correlation matrix to reduce the covariate set. Eigen decomposition is a matrix factorization technique that is useful for characterizing patterns in high-dimensional data (Harville, 1997). Eigen decomposition (or similar orthogonalization techniques) is used widely for data reduction or compression in statistics, signal processing, pattern recognition, remote sensing, and other fields where high-dimensional and/or highly correlated data are encountered (Gunter et al., 2012; Guanter et al., 2013; Huang, Song, Cui, Peng, & Xu, 2014; Wang, Yu, Wang, Deng, & Zhao, 2014). It is common to see PCA performed by decomposing the covariance matrix in the remote sensing sciences (Richards, 2013). We elected to decompose the correlation matrix due to the natural tendency for upper percentile heights to be more
variable than lower percentile heights. Johnson and Wichern (2007) offer a detailed discussion about how standardizing the variables, i.e., using the correlation matrix, reduces the effect of individual covariate variances on principal component loadings. In Section A1, the mathematical procedure for conducting PCA on standardized variables and transforming the gridded data set to the correct basis for prediction is detailed. Criteria for selecting principal components are also presented in Section A1. For the FEF, NIWOT, and GLEES sites, two principal components were retained and three were selected for MEF.

Some grid cell locations contained covariate values that well exceeded the range of values used for model fitting. This is a result of a relatively small set of observed plots compared to the size and heterogeneity of the study area. To avoid potential issues arising from model extrapolation (Perrin, 1904), grid cells with values outside the support of the observed covariate values were removed from the analysis. Water bodies at GLEES and MEF were also excluded from the analysis. Excluded grid cells are colored black in the subsequent maps.

3. Methods

3.1. Modeling framework

Geostatistical settings typically assume at location \( s \in D \subseteq \mathbb{R}^2 \), where \( s \) is a vector of geographic coordinates in domain \( D \), a Gaussian outcome variable \( y(s) \) is modeled using the regression model,

\[
y(s) = \mathbf{x}(s)\beta + \mathbf{x}(s)\mathbf{w}(s) + \epsilon(s).
\]

Here the linear mean structure that accounts for large scale variation in the outcome is composed of a \( p \times 1 \) vector \( \mathbf{x}(s) \), comprising an intercept and spatially referenced covariates, and an associated column vector of regression coefficients \( \beta = (\beta_0, \beta_1, ..., \beta_p - 1) \). The \( \mathbf{x}(s) \) is a \( q \times 1 \) vector that includes the intercept and those covariates from \( \mathbf{x}(s) \) whose impact on the outcome is posed to vary spatially. This space varying impact is captured through the vector of spatial random effects \( \mathbf{w}(s) = (w_1(s), w_2(s), ..., w_q(s))' \). Various sub-models are formed by specifying different combinations of \( \mathbf{x}(s) \) and associated \( \mathbf{w}(s) \). Customarily, \( \epsilon(s) \) is modeled as an independent white-noise process that captures measurement error or micro-scale variation. With any collection of \( n \) locations, say \( S = \{s_1, s_2, ..., s_n\} \), we assume that the \( \epsilon(s) \)'s are independently and identically distributed as \( N(0, \tau^2) \), where \( \tau^2 \) is often called the nugget.

For this model, spatial structure is typically introduced through a multivariate Gaussian process (GP) (Banerjee et al., 2004; Cressie, 1993; Wackernagel, 2003), where a cross-covariance function explicitly models the covariance of \( \mathbf{w}(s) \) both within and among locations. This additional flexibility is attractive in many settings (see, e.g., Banerjee, Finley, Waldmann, and Ericsson (2011); Finley et al. (2011); Gelfand et al. (2004)); however, for the objectives of this study it is not clear if the extra effort will fetch substantial advantages. Rather, for simplicity and substantially reduced computational demand, we assumed the elements of \( \mathbf{w}(s) \) arise from \( q \) independent univariate GPs. Specifically, the process associated with the \( k \)-th covariate is \( w_k(s) - GP(0, \Sigma_k, \theta_k) \) where \( \Sigma_k \) is the \( n \times n \) covariance matrix of \( \mathbf{w}_k \) with \( (i,j) \)-th element given by \( C(s_i, s_j; \theta_k) \). Clearly \( C(s, s'; \theta) \) cannot be just a function; it must ensure that the resulting \( \Sigma_k \) matrix is symmetric and positive definite. Such functions are known as positive definite functions and are characterized as the characteristic function of a symmetric random variable. Further technical details about positive definite functions can be found in Cressie (1993), Chilès and Delfiner (1999), and Banerjee et al. (2004).

We specify \( C(s, s'; \theta_k) = \sigma_k^2 \delta(s - s') + \phi_k(s, s') \) with \( \phi_k = (\phi_k^2, \phi_k^3, ..., \phi_k^p) \) is a valid spatial correlation function, where \( \phi_k \) quantifies the rate of correlation decay and \( \text{Var}(w_k) = \sigma_k^2 \). For the subsequent analyses we assumed an exponential correlation function, \( \rho(||s - s'||; \phi_k) = \exp(-\phi_k ||s - s'||) \), where \( ||s - s'|| \) is the Euclidean distance between the locations \( s \) and \( s' \).

To complete the Bayesian model specification, we assign prior distributions to the model parameters and inference proceeds by sampling from the posterior distribution of the parameters. As customary, we assume \( \beta \) follows a \( \text{MVN}(\mu_0, \Sigma_0) \) prior with \( \mu_0 = 0 \) and \( \Sigma_0 = 10,000 \mathbf{I}_p \), while the spatial variance components \( \sigma_k^2 \) and the measurement error variance \( \tau^2 \) are assigned inverse-Gamma, \( \text{IG}(a, b) \), priors. The process correlation decay parameters \( \phi_k \) follow a Uniform, \( \text{Unif}(a, b) \), with support over the geographic range of the study area.

Using notations similar to Gelman et al. (2013), we can write the posterior distribution of the model parameters as \( p(\Omega | y) \), where \( \Omega = (\beta, \mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_q, \tau^2, \sigma_k^2, \phi_1, \phi_2, ..., \phi_q, \tau^2) \) and \( y = (y(s_1), ..., y(s_n))^' \), which is proportional to

\[
\prod_{k=1}^{q} \text{Unif}(\delta_k; a_k, b_k) \times \prod_{k=1}^{q} \text{IG}(\sigma_k^2; a_k, b_k) \times \text{N}(\mathbf{w}_k | 0, \Sigma_k) \times \text{IG}(\tau^2; a, b) \times \prod_{k=1}^{q} \text{N}(y(s_k) | \mathbf{x}(s_k) \beta + \mathbf{x}(s_k) \mathbf{w}(s_k), \tau^2).
\]

(2)

An efficient MCMC algorithm for estimation of Eq. (2) is obtained by updating \( \beta \) from its full conditional then using Metropolis steps for the remaining parameters. Alternatively, the model can be reparameterized such that the spatial random effects \( \mathbf{w} \) do not need to be sampled directly (Banerjee, Gelfand, Finley, & Huiyan, 2008). In either case, the MCMC algorithms yield posterior samples of \( \Omega \).

For predictions, if \( \mathbf{S}_0 = \{s_{01}, s_{02}, ..., s_{0m}\} \) is a collection of \( m \) new locations, the posterior predictive distribution of the spatial random effects associated with the \( k \)-th regression coefficient is given by

\[
p(w_k | y) = \int p(w_k | w_k, \Omega, y) p(w_k, \Omega, y) p(\Omega, y) d\Omega d\mathbf{w}_k.
\]

(3)

where \( w_k = (w_k(s_{01}), w_k(s_{02}), ..., w_k(s_{0m}))' \).

Because we are using MCMC sample-based inference, i.e., drawing samples from the posterior of \( \Omega \), the integral in Eq. (3) does not have to be evaluated explicitly. Rather, given \( l \) posterior samples for \( \Omega \), i.e., \( \Omega_1, ..., \Omega_l \), this distribution can be obtained via composition sampling (see, e.g., Banerjee et al. (2008) pg. 132) by first draw \( \mathbf{w}^{(l)} \) and then draw \( \mathbf{w}_k^{(l)} \) for each \( l \) from \( p(w_k | w_k, \Omega_l, y) \), where this last distribution is derived as a conditional distribution from a multivariate normal and, hence, is multivariate normal. More precisely, the process realizations over the new locations are conditionally independent of the observed outcomes given the realizations over the observed locations and the process parameters. In other words, \( p(w_k | \mathbf{w}_k, \Omega_l, y) = p(w_k | \mathbf{w}_k, \Omega_l) \), which is a multivariate normal distribution with mean and variance given by

\[
E[w_k | \mathbf{w}_k, \Omega_l] = \text{Cov}(w_k, \mathbf{w}_k) \text{Var}^{-1}(\mathbf{w}_k) w_k = R_0(\mathbf{d}_k) R_0(\mathbf{d}_k)^{-1} w_k
\]

(4)

and \( \text{Var}[w_k | \mathbf{w}_k, \Omega_l] = \sigma_k^2 \left( R_0(\mathbf{d}_k) - R_0(\mathbf{d}_k) R_0(\mathbf{d}_k)^{-1} R_0(\mathbf{d}_k) \right) \),

(5)

where \( R_0(\mathbf{d}_k) \) is the \( n \times m \) matrix with \((i,j)\)-th element given by \( r(s_i, s_j; \theta_k) \) and \( R(\mathbf{d}_k) \) is the \( n \times n \) matrix with \((i,j)\)-th element given by \( r(s_i, s_i; \theta_k) \). This procedure is repeated to generate samples for all \( w_{0k} \)'s. Finally, given a set of covariates at a new location \( s_0 \), samples from the posterior predictive distribution of the outcome variable, \( y(s_0) \), are drawn from \( N(\mathbf{x}(s_0) \beta | \mathbf{x}(s_0) \beta' + \mathbf{x}(s_0) \mathbf{w}^{(l)} , \tau^2) \) for \( l = 1, 2, ..., L \). 95% credible interval widths (CIW) can be calculated using pixel-level posterior predictive distributions of AGB by measuring the distance between the 2.5% and 97.5% quantile breaks (Fig. 2). The 95%
CIWs serve as spatially explicit summaries of prediction uncertainty and can be mapped in order to assess model prediction performance spatially.

3.2. Candidate models

Five candidate models were derived from Eq. (1) and include: non-spatial, where $w_i$'s are set to zero; SVI, where only the spatial random effects associated with the model intercept are included; the full SVC, where all regression coefficients have associated spatial random effects; SVC-$\beta_1$, where spatial random effects for the intercept and $\beta_1$ are included, and; SVC-$\beta_2$, where spatial random effects for the intercept and $\beta_2$ are included. Due to MEF’s additional covariate, four additional candidate models were tested. SVC-$\beta_3$ has spatial random effects associated with $\beta_0$ and $\beta_3$; SVC-$\beta_1\beta_2$ has spatial random effects associated with $\beta_0$, $\beta_1$, and $\beta_2$; SVC-$\beta_1\beta_3$ has spatial random effects associated with $\beta_0$, $\beta_1$, and $\beta_3$; and; SVC-$\beta_2\beta_3$ has spatial random effects associated with $\beta_0$, $\beta_2$, and $\beta_3$.

Empirical semivariograms fit to the residuals of the non-spatial models (Fig. 3) were used to help guide hyperprior specification for the candidate models’ IG and Unif priors. Specifically, for the variance parameters the IG hyperprior $a$ was set equal to 2, which results in a

Fig. 2. Posterior predicted distribution at an example grid cell location shown in gray. Blue vertical dashed lines indicate the 2.5% and 97.5% quantiles of the posterior predicted distribution. The length of the horizontal red line is the 95% credible interval width (CIW). CIWs for each grid cell are mapped in Fig. 5.

Fig. 3. Semivariogram models generated using the residuals from the non-spatial model estimates for FEF, MEF, GLEES, and NIWOT.
prior distribution mean equal to $b$ and infinite variance (see, e.g., IG definition in Gelman et al. (2013)). Then the $b$ hyperpriors for the models' $\tau^2$ and $\phi$'s were set according to the nugget and partial sill of the semivariograms, respectively. The prior for the spatial decay parameters, $\phi$'s, was set to Unif[0.00006, 3] which, assuming the exponential spatial correlation function, corresponds to support for an exponential spatial range between ~1 and 5000 m. The effective spatial range is defined here as the distance at which the correlation equals 0.05.

The MCMC samplers were implemented in C++ and Fortran and leveraged Intel’s Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. All analyses were conducted on a Linux workstation using two Intel Nehalem quad-Xeon processors.

Three MCMC chains were run for 25,000 iterations each. The most demanding model required ~1 h to complete a single MCMC chain. Convergence was diagnosed using the CODA package in R by monitoring mixing of chains and the Gelman–Rubin statistic (Gelman & Rubin, 1992). Satisfactory convergence was diagnosed within 10,000 iterations for all models. Posterior inference was based on a post burn-in sub-sample of 15,000 iterations—every third sample from the last 15,000 iterations of each chain.

### 3.3. Fit and prediction accuracy assessment

Model performance was assessed using the popular Deviance Information Criterion (DIC) to rank models in terms of how well they fit the data (Spiegelhalter, Best, Carlin, & van der Linde, 2002). This criteria is the sum of the Bayesian deviance (a measure of model goodness of fit) and the effective number of parameters (a penalty for model complexity), $D$ and $pD$, respectively. Lower values of DIC indicate better model fit.

### Table 1
Candidate model parameter estimates and 95% credible intervals for FEF.

<table>
<thead>
<tr>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_\beta$</td>
</tr>
<tr>
<td>$\tau^2$</td>
</tr>
<tr>
<td>$\phi_i$</td>
</tr>
<tr>
<td>$\rho_i$</td>
</tr>
</tbody>
</table>

### Table 2
Candidate model parameter estimates and 95% credible intervals for MEF.

<table>
<thead>
<tr>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_\beta$</td>
</tr>
<tr>
<td>$\tau^2$</td>
</tr>
<tr>
<td>$\phi_i$</td>
</tr>
<tr>
<td>$\rho_i$</td>
</tr>
</tbody>
</table>

---


---

118

10.1016/j.rse.2015.01.012
Table 3
Candidate model parameter estimates and 95% credible intervals for GLEES.

<table>
<thead>
<tr>
<th>Parameter C.I.</th>
<th>Non-spatial</th>
<th>SVI</th>
<th>SVC-(\phi_1)</th>
<th>SVC-(\phi_2)</th>
<th>SVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(50% (2.5%, 97.5%))</td>
<td>(\beta_0)</td>
<td>8.27 (7.66, 8.86)</td>
<td>8.40 (7.76, 9.14)</td>
<td>8.31 (7.61, 9.03)</td>
<td>7.94 (7.43, 8.51)</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.66 (0.37, 0.96)</td>
<td>0.67 (0.37, 0.99)</td>
<td>0.86 (0.49, 1.23)</td>
<td>0.74 (0.14, 1.29)</td>
<td>0.87 (0.33, 1.45)</td>
</tr>
<tr>
<td>(\tau)</td>
<td>4.10 (2.75, 6.63)</td>
<td>0.53 (0.19, 2.01)</td>
<td>0.40 (0.10, 1.67)</td>
<td>0.32 (0.10, 1.22)</td>
<td>0.25 (0.08, 0.95)</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>44.11 (12.90, 173.16)</td>
<td>33.88 (10.46, 327.24)</td>
<td>22.85 (10.31, 125.63)</td>
<td>20.96 (10.27, 151.98)</td>
<td>20.19 (10.27, 198.41)</td>
</tr>
<tr>
<td>(\omega_2)</td>
<td>3.45 (1.95, 5.88)</td>
<td>0.98 (0.35, 2.66)</td>
<td>0.71 (0.29, 1.77)</td>
<td>0.53 (0.24, 1.33)</td>
<td>0.39 (0.21, 0.85)</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>4.01 (2.82, 5.98)</td>
<td>1.29 (0.36, 3.74)</td>
<td>1.09 (0.25, 2.76)</td>
<td>0.99 (0.21, 3.14)</td>
<td>0.66 (0.20, 2.10)</td>
</tr>
<tr>
<td>(\psi)</td>
<td>116.58 (13.64, 1444.64)</td>
<td>60.19 (10.52, 1578.95)</td>
<td>137.24 (11.40, 1167.32)</td>
<td>42.98 (10.43, 1401.87)</td>
<td>77.34 (10.46, 2419.35)</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>2.66 (0.78, 5.07)</td>
<td>1.19 (0.34, 3.21)</td>
<td>1.74 (0.41, 3.87)</td>
<td>0.87 (0.30, 2.51)</td>
<td>0.38 (0.20, 1.00)</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>0.21 (0.04, 0.49)</td>
<td>0.21 (0.05, 0.48)</td>
<td>0.31 (0.01, 0.62)</td>
<td>0.23 (0.52, 0.78)</td>
<td>0.34 (0.17, 0.82)</td>
</tr>
<tr>
<td>(\sigma_0^2)</td>
<td>3.17 (2.75, 6.63)</td>
<td>0.53 (0.19, 2.01)</td>
<td>0.40 (0.10, 1.67)</td>
<td>0.32 (0.10, 1.22)</td>
<td>0.25 (0.08, 0.95)</td>
</tr>
<tr>
<td>(\sigma_1^2)</td>
<td>3.45 (1.95, 5.88)</td>
<td>0.98 (0.35, 2.66)</td>
<td>0.71 (0.29, 1.77)</td>
<td>0.53 (0.24, 1.33)</td>
<td>0.39 (0.21, 0.85)</td>
</tr>
<tr>
<td>(\sigma_2^2)</td>
<td>4.01 (2.82, 5.98)</td>
<td>1.29 (0.36, 3.74)</td>
<td>1.09 (0.25, 2.76)</td>
<td>0.99 (0.21, 3.14)</td>
<td>0.66 (0.20, 2.10)</td>
</tr>
</tbody>
</table>

4. Results

The experimental design produced 24 combinations of candidate models and sites. Summaries of the candidate models’ parameter estimates, fit, and predictive performance for FEF, MEF, GLEES and NIWOT are provided in Tables 1, 2, 3, and 4. Concerning all sites’ non-spatial models, the exclusion of zero from the regression slope parameters’ 95% credible interval suggested the derived LiDAR covariates explained a substantial portion of variability in AGB, with the exception of NIWOT’s parameters. Non-spatial SVI SVC-\(\phi_1\) SVC-\(\phi_2\) SVC

<table>
<thead>
<tr>
<th>Fit</th>
<th>Non-spatial</th>
<th>SVI</th>
<th>SVC-(\phi_1)</th>
<th>SVC-(\phi_2)</th>
<th>SVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>196.30</td>
<td>103.69</td>
<td>90.35</td>
<td>79.59</td>
<td>67.79</td>
</tr>
<tr>
<td>DIC</td>
<td>200.20</td>
<td>133.48</td>
<td>117.93</td>
<td>109.85</td>
<td>101.53</td>
</tr>
<tr>
<td>RMSPE</td>
<td>35.79</td>
<td>34.21</td>
<td>35.63</td>
<td>35.02</td>
<td>35.20</td>
</tr>
</tbody>
</table>

Predictive performance was assessed using a 10-fold cross-validation approach. Here, the full dataset was split into ten roughly equal sized subsets. Then AGB was predicted for locations within each subset, given parameters estimated from the remaining subsets. Root mean squared prediction error (RMSPE) was then calculated using the observed AGB values and corresponding median of the posterior predictive distribution. Lower RMSPE values signify more accurate predictions.
The remainder of this section is divided into four subsections, one for each site. In each subsection the “best” model is identified and some site-specific results are explored.

4.1. FEF results

At FEF, the full SVC model produced the lowest $D_{IC}$, and $RMSPE$. The SVC model showed an 83% improvement in $DIC$ compared to the non-spatial model. There was high correspondence between the observed and fitted values produced with the SVC model, whereas the fitted values for the non-spatial model were scattered further from the plot data (Fig. A2). The credible intervals for the SVC fitted values were also much tighter than the non-spatial fitted values (Fig. A2). There was a 9% improvement in $RMSPE$ when moving from the non-spatial to SVC model. Most of the predictions produced with the SVC model fell within the range of observed AGB (0–430 Mg/ha, Fig. 4a). Those pixels that exceeded this observed range exhibited high associated uncertainties (Fig. 5a). For example, SVC model predictions in a small patch in the southeastern corner of the FEF domain showed large uncertainty, perhaps indicating that forest structure characteristics in the region are not adequately represented in the plot data used to estimate model parameters. Further, there is a pattern of large uncertainty across the northern region of the FEF, again suggesting a paucity of plot data to characterize the relationship between AGB and the LiDAR covariates.

The mean, or median, posterior predictive surface of $\beta_k(s) = \beta_k + w_k(s)$ can provide insights into the space-varying relationship between the $k$-th covariate and AGB. For example, the $\beta_1(s)$ coefficient map showed clear spatial structure and suggested that the associated LiDAR covariate explained more variability in AGB in the northern region of the study area, i.e., $\beta_1(s)$ was farther from zero in the north (Fig. 6b). After examining the LiDAR derived digital terrain model, it could be seen that $\beta_1(s)$ was related to elevation (Fig. 7). This suggests that the relationship between AGB and the LiDAR covariate associated with $\beta_1(s)$ changes as a function of elevation—due perhaps to species composition change or a fertility gradient that influences the relationship between the covariate and stem biomass. Although not further explored here, one could envision adapting this model to include elevation as a covariate to potentially eliminate the need for $\beta_1$ to vary spatially.

Compared to $\beta_1(s)$, random effects associated with $\beta_0(s)$ and $\beta_2(s)$ had much shorter effective ranges (Table 1), which made their adjustments less noticeable at the scale shown in Fig. 6. By zooming into a region with a high concentration of sub-plots, switching from an equal to quantile interval color classification, and passing the predictions of the $\beta(s)$’s through a smoothing filter, the more subtle spatial adjustments to $\beta_0$ and $\beta_2$ could be seen. Fig. 8a and b depicts $\beta_0(s)$ and $\beta_2(s)$ at the extents shown by the red outline in Fig. 6a and c, respectively. The zoomed-in figures show that sub-plot clusters received slightly higher or lower $\beta_0(s)$ and $\beta_2(s)$, confirming that the random effects are capturing local or micro-scale variability. The global estimates of $\beta_0$ and $\beta_2$ (~11.28 and ~0.91 respectively) are influenced by their corresponding spatial random effects to better accommodate the non-stationary relationship between the covariates and AGB.

4.2. MEF results

As noted previously, of the eight spatial models fit to the MEF dataset none outperformed the non-spatial model with regard to prediction...
accuracy. Regarding fit to the observed data, only the SVI model showed an improvement in DIC over the non-spatial model (Table 2).

Results from MEF provided an example of how uncertainty maps can aid in identifying spatial extrapolation issues. Figs. 4b and 5b show the non-spatial model AGB prediction and associated uncertainty for the extent of the LiDAR data. A large number of grid cells in the southwest corner of MEF were removed from the prediction dataset because their covariate values far exceeded those seen in the observed dataset (see Section 2.6). Also, grid cells in this area that were included in the prediction set had high uncertainties. An examination of MEF’s aerial photo suggested that the forest in this region has different characteristics than the areas where plot data were collected (Fig. 1b). In this instance, having uncertainty maps for MEF helped to diagnose a spatial extrapolation issue that may have otherwise gone unnoticed.

4.3. GLEES results

The SVI model at GLEES had the lowest RMSPE value of the five models (Table 3); however, all three SVC variant models had lower D and DIC values. Here, the increased flexibility afforded by the additional spatial random effects in the SVC models improved the fit but may have reduced prediction accuracy due to over-fitting. The SVI model exhibited a 33% improvement in DIC and a 4.5% increase in prediction accuracy over the non-spatial model. Because prediction is our focus, the SVI model was preferred for the GLEES dataset (Figs. 4c and 5c).

4.4. NIWOT results

The SVC-β2 model provided the most accurate prediction for the NIWOT dataset. This model produced a 24% lower DIC and 7% lower RMSPE value compared to the non-spatial model. As with GLEES, the more complex SVC model provided better fit (5% lower than SVC-β2), but this did not translate into improved prediction accuracy.

NIWOT’s non-spatial model β2 parameter was statistically not significant, i.e., zero was contained in its 95% credible interval. In traditional regression model selection procedures, parameters deemed not significant would typically be removed. Here, even though β2 was globally not significant, allowing the parameter to vary across the spatial domain improved model fit and predictive performance. It can be seen that some clusters of sub-plots received higher estimated values of β2(s) than others (Fig. 10b). This result provided evidence that in some locations, the LiDAR covariate associated with β2 was useful for explaining variability in AGB.

5. Discussion

The usefulness of Bayesian space-varying coefficient models for predicting AGB was assessed at four forested NACP sites. For three of the four sites, the non-spatial model residuals exhibited enough spatial correlation to warrant the use of a spatial random effect on the intercept. For these sites, the presence of serial correlations among the residuals suggested that the assumptions of the non-spatial model were violated, making the validity of any subsequent inference based on
parameter estimates or prediction maps using the non-spatial model questionable. Further, beyond meeting model assumptions, explicitly accommodating the residual dependence via spatially structured random effects fetched improved prediction by borrowing information from proximate observed locations. This was illustrated in the analysis of the FEF, GLEES, and NIWOT datasets, where the SVI models marginally outperformed the non-spatial models (i.e., providing lower RMSPE). A potential reason for seeing only slight prediction accuracy improvements may have been due to the small number of inventory plots. With such small sample sizes it is not possible to obtain precise estimates of spatial covariance parameters. Hence, we might expect some impact on prediction accuracy when the covariance uncertainty is propagated through to the posterior predictive distributions. Also, the sparsity of inventory plots means that there is little information to borrow from when making predictions—this is exacerbated by cross-validation, which further reduces the sample size.

For the FEF and NIWOT datasets, allowing regression coefficients associated with LiDAR covariates to vary spatially improved prediction accuracy beyond that achieved by the SVI models (i.e., the SVC models showed lower RMSPE). Although the improvements in RMSPE were small, these results suggested the more complex SVC models were able to estimate and leverage non-stationary spatial relationships between AGB and the LiDAR covariates. There was no advantage, from a prediction standpoint, for SVC models at MEF or GLEES. Given the small number of inventory plots and relatively homogeneous forest characteristics within the NACP sites, it is not too surprising that the SVC models did not perform consistently better than the simpler models.

In studies based on larger datasets distributed across more heterogeneous domains, substantial improvements in prediction accuracy have been obtained by adding spatial random effects to the model intercept, see, e.g., Babcock, Matney, Finley, Weiskittel, and Cook (2013), and

Fig. 6. Maps showing the spatially varying coefficients and associated 95% credible interval width (CIW) for the SVC model at FEF. Black squares indicate excluded prediction grid cells (see Section 2.6).
regression coefficients, see, e.g., Finley et al. (2009), Finley et al. (2011), Jarzyna et al. (2014). In these studies, the SVC models were effective because there were strong non-stationary relationships between the covariates and outcome variable—driven by unobserved environmental factors that operated at broad spatial scales. Also, substantially larger sample sizes \((n > 1000\) in all cases) allowed more precise estimation of the spatial covariance parameters that, in turn, improved the posterior predictive distribution precision and perhaps accuracy.

Study results underscored the difference between goodness of fit metrics, e.g., DIC, and out of set prediction performance, assessed here via cross-validation. In three of the four datasets, the exception being FEF, DIC favored a more complex model over the simpler model selected based on RMSPE. The flexibility added by allowing the intercept and regression slope coefficients to vary spatially will generally allow fitted values to more closely approximate the observed data. Unlike an unstructured random effect model, see, e.g., Salas, Ene, Gregoire, Næsset, and Gobakken (2010), the Gaussian Process imposes some smoothing that should reduce over-fitting to the observed data. However, the flexibility afforded by a univariate Gaussian Process on each

---

**Fig. 7.** Digital terrain model for FEF.

**Fig. 8.** Maps showing the spatially varying coefficients zoomed into the extent outlined in red in Fig. 6a and 6c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green. Black squares indicate excluded prediction grid cells (see Section 2.6).

**Fig. 9.** Maps showing the spatially varying coefficients and associated 95% credible interval width (CIW) for the SVC-β2 model at NIWOT. Black squares indicate excluded prediction grid cells (see Section 2.6).
regression coefficient appears to result in some over-fitting and hence reduced prediction performance. Because prediction is the primary goal of many AGB modeling studies it is important to base model selection on metrics that summarize out of sample predictions, e.g., RMSPE, and not goodness of fit.

Uncertainty quantification is an important component of carbon Measurement, Reporting, and Verification systems (MRVs) like those called for by the United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries and actively being developed by NASA’s Carbon Monitoring System Science Team (CMS, 2010; UN-REDD, 2009). Maps like the ones developed in Fig. 5 can aid uncertainty analysis by providing a tool to assess the reliability of AGB predictions. As shown with the FEF and MEF results, spatially explicit representations of uncertainty can be used to identify regions of concern regarding spatial extrapolation and highlight areas where more field data may be required.

Mapping space-varying coefficients can shed light on missing covariates. At FEF, examining how the $\beta_1(s)$ parameter varied spatially revealed that elevation may be a missing covariate in the AGB models. If further AGB model refinement was pursued at FEF, the inclusion of elevation as a covariate, either in an additive or multiplicative fashion, is warranted.

Bayesian space-varying coefficient models provide statistically valid inference about non-stationary relationships between the covariates and the outcome variable. Results from the NIWOT analysis showed this phenomenon when $\beta_2$ was allowed to vary spatially. Examining regression parameters’ spatial impact via the proposed modeling framework can expose locally important covariates, improve overall prediction accuracy, and potentially provide insight into ecological processes.

Given the small size of the datasets considered here, the proposed modeling framework was computationally feasible. However, when the datasets consist of several thousand observations, which is common in forest inventory databases, cubic order matrix operations required for evaluating the model likelihood make parameter estimation computationally onerous. Therefore, our future work will focus on exploring strategies for dimension reduction when fitting SVC models. Further, we will extend the SVC models to accommodate multiple forest variables of interest.

Acknowledgments

This work was partially supported by grants from the National Science Foundation DMS-1106609, EF-1137309, EF-1241874, and EF-1253225, National Aeronautical and Space Administration Carbon Monitoring System and NNI08AH971, and the USDA Forest Service. Special thanks to the anonymous reviewers.

Appendix A

A1. Principal components analysis procedure

Let $A$ be the sub-plot and $B$ be the grid percentile height sets with observations along the rows and percentile heights along the columns. Their dimensions are $n \times p$ and $m \times p$, respectively, where $n$ is the number of observations (i.e., sub-plot count), $m$ is the number of prediction cells, and $p$ is the number of percentile height covariates ($p = 20$). Let $\tilde{A}$ be the standardized scores for $A$ calculated as follows,

$$\tilde{A}_{ij} = \frac{A_{ij} - \mu_j}{\sigma_j},$$

(A1)

where $\mu_j$ and $\sigma_j$ are the mean and standard deviation of the $j$-th column of $A$. A correlation matrix $\rho$ is constructed for $A$ by,

$$\rho = \frac{\tilde{A} A^*}{n-1}.$$  

(A2)

Let $\rho = P A P^T$ be the eigen decomposition of $\rho$ with the diagonal elements of $A$ ($p \times p$) holding the eigenvalues of $\rho$ in decreasing order (off diagonal elements are zero) and $P$ ($p \times p$) holding the corresponding eigenvectors along the columns. $P$ is an orthogonal transformation matrix that can be used to project $\tilde{A}$ into a new vector space. Let $A^* = \tilde{A} P$ be the transformed standardized scores of the sub-plot percentile height matrix. The columns of $A^*$ are now uncorrelated.

To transform the gridded data $B$ to the same vector space as $A^*$, $B$ was first standardized via,

$$\tilde{B}_{ij} = \frac{B_{ij} - \mu_j}{\sigma_j}.$$  

(A3)

Note that $\mu_j$ and $\sigma_j$ are the mean and standard deviation calculated from $A$. Then $B^* = BP$ is then the gridded percentile height data projected into the same vector space as $A^*$.

In practice, a subset of columns from $A^*$ (i.e., principal component scores) can be used as covariates in regression analysis—referred to as principal components regression (Chatterjee & Hadi, 2006). This is the approach we pursued here. The columns of $A^*$ represent a new, orthogonal (i.e., uncorrelated), candidate covariate set. From this set we retained the minimum number of covariates that explained at least 80% of the variance in the percentile height data. This was done by selecting the leftmost columns of $A^*$ whose standardized eigenvalues total was $>0.8$. To meet this criterion the two leftmost columns of $A^*$ were retained for FEF, NIWOT and GLEES. It was necessary to keep three columns for MEF. Fig. A1 shows a plot of the cumulative sum of the eigenvalues for each site. We kept the corresponding columns of $B^*$ for each site to use in subsequent grid prediction and mapping.
Fig. A1. The cumulative sums of the standardized eigenvalues (in decreasing order) for FEF, MEF, GLEES, and NIWOT.

Fig. A2. Plots showing fitted versus field-observed AGB values for FEF’s non-spatial and SVC models. Vertical Gray error bars depict 95% credible intervals for fitted values.
Fig. A3. Plots showing fitted versus field-observed AGB values for MEF’s non-spatial and SVI models. Vertical gray error bars depict 95% credible intervals for fitted values.

Fig. A4. Plots showing fitted versus field-observed AGB values for GLEES’s non-spatial and SVI models. Vertical gray error bars depict 95% credible intervals for fitted values.

Fig. A5. Plots showing fitted versus field-observed AGB values for NIWOT’s non-spatial and SVC-β2 models. Vertical gray error bars depict 95% credible intervals for fitted values.