Remote Sensing of Environment 169 (2015) 113–127

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

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

ARTICLE INFO

Article history:
Received 9 April 2014
Received in revised form 4 June 2015
Accepted 28 July 2015
Available online xxxx

Keywords:
Bayesian hierarchical models
Markov chain Monte Carlo
Gaussian process
Geospatial
LiDAR
Forest biomass

ABSTRACT

Many studies and production inventory systems have shown the utility of coupling covariates derived from Light Detection and Ranging (LiDAR) data with forest variables measured on georeferenced inventory plots through regression models. The objective of this study was to propose and assess the use 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. We explored this objective using four forest inventory datasets that are part of the North American Carbon Program, each comprising point-referenced measures of above-ground forest biomass and discrete LiDAR. For each dataset, we considered at least five regression model specifications of varying complexity. Models were assessed based on goodness of fit criteria and predictive performance using a 10-fold cross-validation procedure. Results showed that the addition of spatial random effects to the regression model intercept improved fit and predictive performance in the presence of substantial residual spatial dependence. Additionally, in some cases, allowing either some or all regression slope parameters to vary spatially, via the addition of spatial random effects, further improved model fit and predictive performance. In other instances, models showed improved fit but decreased predictive performance—indicating over-fitting and underscoring the need for cross-validation to assess predictive ability. The proposed Bayesian modeling framework provided access to pixel-level posterior predictive distributions that were useful for uncertainty mapping, diagnosing spatial extrapolation issues, revealing missing model covariates, and discovering locally significant parameters.

© 2015 Elsevier Inc. All rights reserved.

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 attributions. 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.

* Corresponding author.
E-mail address: finleya@msu.edu (A.O. Finley).

http://dx.doi.org/10.1016/j.rse.2015.07.028
0034-4257/© 2015 Elsevier Inc. All rights reserved.

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.

SVI and SVC 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 map 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 *Populus tremuloides* Michx. and *grandidentata* 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 (Guanter 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

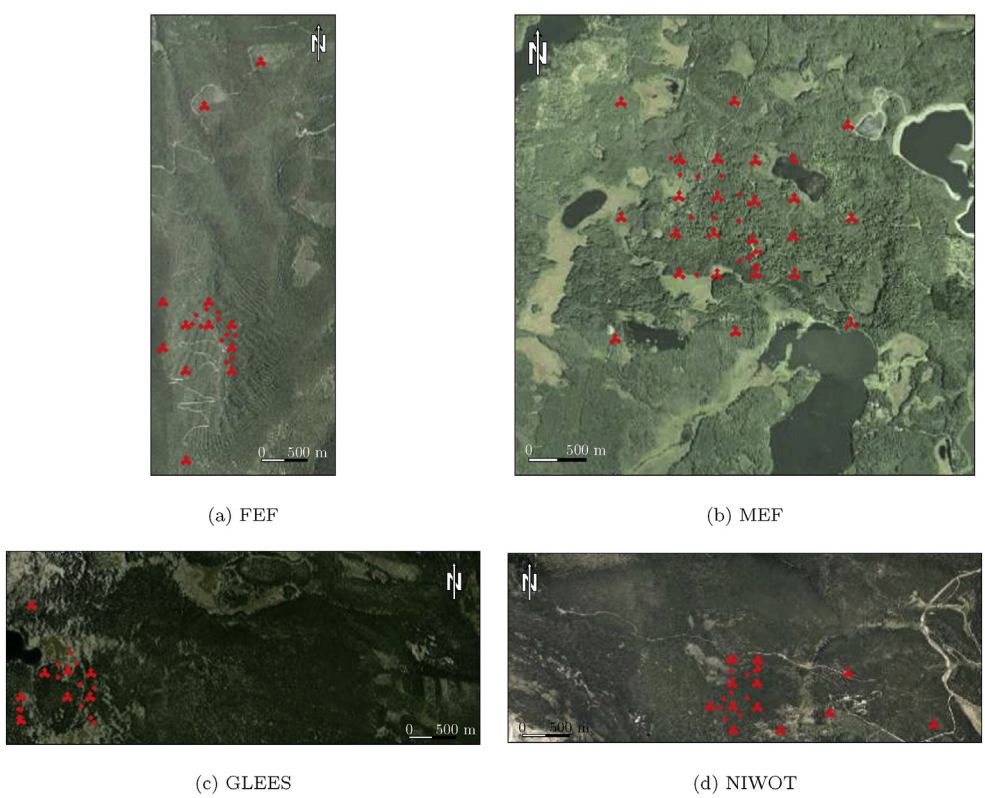


Fig. 1. Aerial photographs of FEF, MEF, GLEES, and NIWOT with inventory sub-plot locations highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 $\boldsymbol{s} \in D \subseteq \mathbb{R}^2$, where $\boldsymbol{s}$ is a vector of geographic coordinates in domain $\mathcal{D}$, a Gaussian outcome variable $y(\boldsymbol{s})$ is modeled using the regression model,

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta} + \tilde{\boldsymbol{x}}(\boldsymbol{s})'\boldsymbol{w}(\boldsymbol{s}) + \epsilon(\boldsymbol{s}). \tag{1}$$

Here the linear mean structure that accounts for large scale variation in the outcome is composed of a $p \times 1$ vector $\boldsymbol{x}(\boldsymbol{s})$, comprising an intercept and spatially referenced covariates, and an associated column vector of regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$. The $\tilde{\boldsymbol{x}}(\boldsymbol{s})$ is a $q \times 1$ vector that includes the intercept and those covariates from $\boldsymbol{x}(\boldsymbol{s})$ whose impact on the outcome is posited to vary spatially. This space varying impact is captured through the vector of spatial random effects $\boldsymbol{w}(\boldsymbol{s}) = (w_1(\boldsymbol{s}), w_2(\boldsymbol{s}), \ldots, w_q(\boldsymbol{s}))'$. Various sub-models are formed by specifying different combinations of $\tilde{\boldsymbol{x}}(\boldsymbol{s})$ and associated $\boldsymbol{w}(\boldsymbol{s})$. Customarily, $\epsilon(\boldsymbol{s})$ is modeled as an independent white-noise process that captures measurement error or micro-scale variation. With any collection of n locations, say $\mathcal{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$, we assume that the $\epsilon(\boldsymbol{s}_i)$'s are identically and independently distributed as $N(0, \tau^2)$, where τ^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 $\boldsymbol{w}(\boldsymbol{s})$ 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 $\boldsymbol{w}(\boldsymbol{s})$ arise from q independent univariate GPs. Specifically, the process associated with the k-th covariate is $w_k(\boldsymbol{s}) \sim GP(0, C(\cdot, \cdot; \boldsymbol{\theta}_k))$ where $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k) = Cov\,(w_k(\boldsymbol{s}), w_k(\boldsymbol{s}^*))$ is a valid covariance function that models the covariance corresponding to a pair of locations $\boldsymbol{s}$ and $\boldsymbol{s}^*$. The process realizations are collected into an $n \times 1$ vector, say $\boldsymbol{w}_k = (w_k(\boldsymbol{s}_1), \ldots, w_k(\boldsymbol{s}_n))'$, that follows a multivariate normal distribution $MVN(0, \Sigma_k)$, where Σ_k is the $n \times n$ covariance matrix of $\boldsymbol{w}_k$ with (i, j)-th element given by $C(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}_k)$. Clearly $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k)$ cannot be just any function; it must ensure that the resulting Σ_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 $C(\boldsymbol{s}, \boldsymbol{s}^*; \boldsymbol{\theta}_k) = \sigma_k^2 \rho(\boldsymbol{s}, \boldsymbol{s}^*; \phi_k)$ with $\boldsymbol{\theta}_k = \{\sigma_k^2, \phi_k\}$, $\rho(\cdot; \phi_k)$ is a valid spatial correlation function, where ϕ_k quantifies the rate of correlation decay and $\mathrm{Var}(w_k) = \sigma_k^2$. For the subsequent analyses we assumed an exponential correlation function, $\rho(\| \boldsymbol{s} - \boldsymbol{s}^* \|; \phi_k) = \exp(-\phi_k \| \boldsymbol{s} - \boldsymbol{s}^* \|)$, where $\| \boldsymbol{s} - \boldsymbol{s}^* \|$ is the Euclidean distance between the locations $\boldsymbol{s}$ and $\boldsymbol{s}^*$.

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 $\boldsymbol{\beta}$ follows a $MVN(\boldsymbol{\mu}_\beta, \Sigma_\beta)$ prior with $\boldsymbol{\mu}_\beta = 0$ and $\Sigma_\beta = 10,000\boldsymbol{I}_p$, while the spatial variance components σ_k^2's and the measurement error variance τ^2 are assigned inverse-Gamma, $IG(a, b)$, priors. The process correlation decay parameters ϕ_k's follow a Uniform, $Unif(a, b)$, 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 $p(\Omega \mid \boldsymbol{y})$, where $\Omega = \{\boldsymbol{\beta}, \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q, \sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2, \phi_1, \phi_2, \ldots, \phi_q, \tau^2\}$ and $\boldsymbol{y} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))'$, which is proportional to

$$\prod_{k=1}^{q} Unif(\phi k|a_{\phi_k}, b_{\phi_k}) \times \prod_{k=1}^{q} IG(\sigma_k^2|a_{\sigma_k}, b_{\sigma_k}) \times N\left(\boldsymbol{\beta}\middle|\boldsymbol{\mu}_\beta, \Sigma_\beta\right) \times IG(\tau^2|a_\tau, b_\tau) \times \\ \times \prod_{k=1}^{q} N(\boldsymbol{w}_k|\boldsymbol{0}, \Sigma_k) \times \prod_{i=1}^{n} N\left(y(\boldsymbol{s}_i)|\boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta} + \tilde{\boldsymbol{x}}(\boldsymbol{s})'\boldsymbol{w}(\boldsymbol{s}), \tau^2\right). \tag{2}$$

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

For predictions, if $S_0 = \{\boldsymbol{s}_{0,1}, \boldsymbol{s}_{0,2}, \ldots, \boldsymbol{s}_{0,m}\}$ 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

$$p(\boldsymbol{w}_{k,0}|\boldsymbol{y}) \propto \int p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega, \boldsymbol{y}) p(\boldsymbol{w}_k|\Omega, \boldsymbol{y}) p(\Omega|\boldsymbol{y}) d\Omega \boldsymbol{w}_k, \tag{3}$$

where $\boldsymbol{w}_{k,0} = (w_k(\boldsymbol{s}_{0,1}), w_k(\boldsymbol{s}_{0,2}), \ldots, w_k(\boldsymbol{s}_{0,m}))'$.

Because we are using MCMC sample-based inference, i.e., drawing samples from the posterior of Ω, the integral in Eq. (3) does not have to be evaluated explicitly. Rather, given L posterior samples for Ω, i.e., $\{\Omega^{(l)}\}_{l=1}^{L}$, this distribution can be obtained via composition sampling (see, e.g., Banerjee et al. (2008) pg. 132) by first draw $\boldsymbol{w}_k^{(l)}$ and then draw $\boldsymbol{w}_{k,0}^{(l)}$ for each l from $p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k^{(l)}, \Omega^{(l)}, \boldsymbol{y})$, 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, $p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega, \boldsymbol{y}) = p(\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega)$, which is a multivariate normal distribution with mean and variance given by

$$\mathrm{E}\left[\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega\right] = \mathrm{Cov}(\boldsymbol{w}_{k,0}, \boldsymbol{w}_k)\mathrm{Var}^{-1}(\boldsymbol{w}_k)\boldsymbol{w}_k = R_0(\phi_k)'R(\phi_k)^{-1}\boldsymbol{w}_k \tag{4}$$

$$\text{and } \mathrm{Var}\left[\boldsymbol{w}_{k,0}|\boldsymbol{w}_k, \Omega\right] = \sigma_k^2\left\{R(\phi_k) - R_0(\phi_k)'R(\phi_k)^{-1}R_0(\phi_k)\right\}, \tag{5}$$

where $R_0(\phi_k)$ is the $n \times m$ matrix with (i, j)-th element given by $\rho(\boldsymbol{s}_{0,i}, \boldsymbol{s}_j; \phi_k)$ and $R(\phi_k)$ is the $n \times n$ matrix with (i, j)-th element given by $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \phi_k)$. This procedure is repeated to generate samples for all $\boldsymbol{w}_{k,0}$'s. Finally, given a set of covariates at a new location $\boldsymbol{s}_0$, samples from the posterior predictive distribution of the outcome variable, $y(\boldsymbol{s}_0)^{(l)}$, are drawn from $N(\boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta}^{(l)} + \tilde{\boldsymbol{x}}(\boldsymbol{s}_0)'\boldsymbol{w}_0^{(l)}, \tau^{2(l)})$ for $l = 1, 2, \ldots, L$.

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%

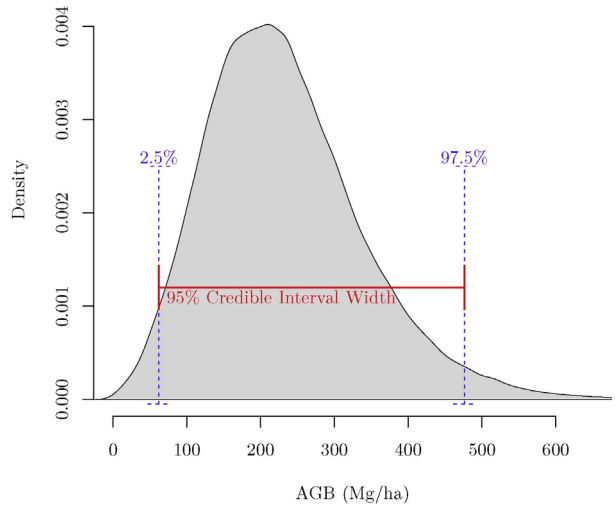


Fig. 2. Posterior predicted distribution at an example grid cell location shown in gray. Blue vertical dashed lines indicate the 2.5% and 97.5% quantiles of the posterior predicted distribution. The length of the horizontal red line is the 95% credible interval width (CIW). CIWs for each grid cell are mapped in Fig. 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 $\boldsymbol{w}_k$'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*-β_1, where spatial random effects for the intercept and β_1 are included, and; *SVC*-β_2, where spatial random effects for the intercept and β_2 are included. Due to MEF's additional covariate, four additional candidate models were tested. *SVC*-β_3 has spatial random effects associated with β_0 and β_3; *SVC*-$\beta_1\beta_2$ has spatial random effects associated with β_0, β_1, and β_2; *SVC*-$\beta_1\beta_3$ has spatial random effects associated with β_0, β_1, and β_3, and; *SVC*-$\beta_2\beta_3$ has spatial random effects associated with β_0, β_2, and β_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

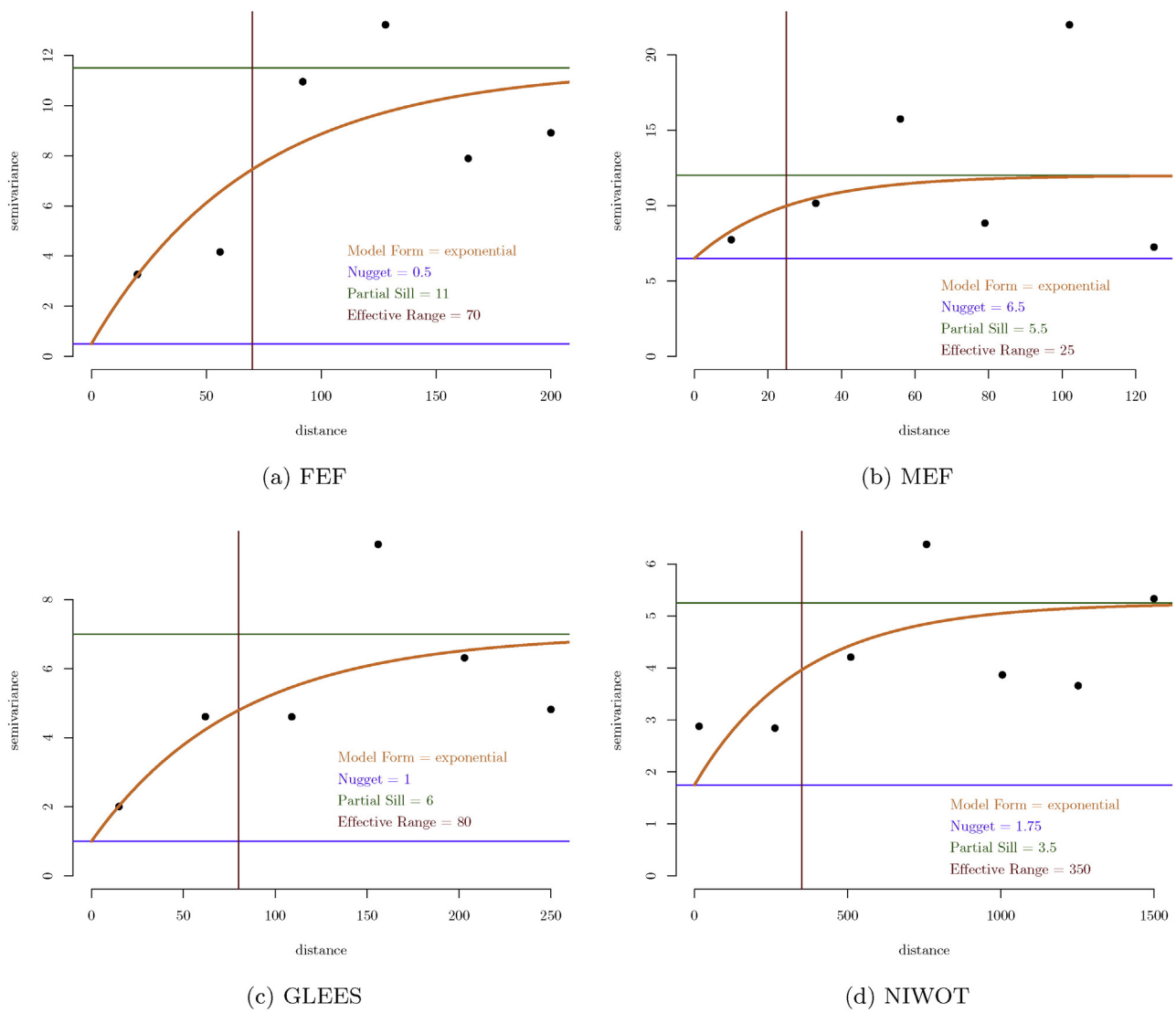


Fig. 3. Semivariogram models generated using the residuals from the *non-spatial* model estimates for FEF, MEF, GLEES, and NIWOT.

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

		Non-spatial	SVI	SVC-β_1	SVC-β_2	SVC
Parameter C.I. 50% (2.5%, 97.5%)	β_0	11.16 (10.49, 11.80)	11.33 (10.44, 12.27)	11.58 (10.88, 12.30)	11.41 (10.69, 12.05)	11.28 (10.69, 11.90)
	β_1	−1.08 (−1.25, −0.92)	−1.06 (−1.23, −0.88)	−1.04 (−1.91, 0.17)	−0.97 (−1.16, −0.78)	−0.81 (−1.76, 0.12)
	β_2	1.10 (0.70, 1.51)	1.01 (0.57, 1.44)	0.77 (0.35, 1.22)	1.46 (0.63, 3.51)	0.91 (0.35, 1.41)
	τ^2	6.49 (4.62, 9.58)	0.89 (0.32, 3.70)	3.38 (1.61, 5.70)	2.36 (0.79, 4.83)	0.07 (0.02, 0.75)
	$3/\phi_0$	–	121.45 (20.05, 284.91)	21.14 (10.29, 241.36)	31.29 (10.37, 1612.90)	20.62 (10.27, 156.90)
	$3/\phi_1$	–	–	2380.95 (160.09, 4838.71)	–	2400.00 (302.72, 4918.03)
	$3/\phi_2$	–	–	–	194.74 (10.82, 2542.37)	21.63 (10.25, 185.87)
	σ_0^2	–	5.18 (2.51, 8.71)	0.48 (0.21, 1.73)	0.53 (0.21, 2.19)	0.48 (0.20, 1.19)
	σ_1^2	–	–	0.66 (0.29, 1.91)	–	0.74 (0.33, 1.95)
	σ_2^2	–	–	–	2.28 (0.80, 7.99)	1.64 (0.89, 3.09)
Fit statistics	D	293.21	93.13	248.42	224.71	22.43
	pD	3.93	38.24	24.17	29.57	25.71
	DIC	297.14	131.37	272.59	254.28	48.14
	RMSPE	71.70	69.44	71.16	74.88	65.27

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' τ^2 and σ^2's were set according to the nugget and partial sill of the semivariograms, respectively. The prior for the spatial decay parameters, ϕ's, was set to *Unif*(0.0006, 3) which, assuming the exponential spatial correlation function, corresponds to support for an *effective 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 subsample 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 p_D, respectively. Lower values of *DIC* indicate better model fit.

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

		Non-spatial	Spatial	SVC-β_1	SVC-β_2	SVC-β_3	SVC-$\beta_1\beta_2$	SVC-$\beta_1\beta_3$	SVC-$\beta_2\beta_3$	SVC
Parameter C.I. 50% (2.5%, 97.5%)	β_0	9.10 (8.45, 9.73)	9.08 (8.43, 9.73)	9.00 (8.30, 9.69)	9.12 (8.44, 9.80)	9.09 (8.45, 9.73)	9.01 (8.27, 9.75)	8.98 (8.27, 9.69)	9.09 (8.39, 9.79)	9.00 (8.22, 9.75)
	β_1	−1.29 (−1.49, −1.09)	−1.29 (−1.48, −1.09)	−1.23 (−1.50, −0.93)	−1.28 (−1.51, −1.06)	−1.28 (−1.49, −1.06)	−1.22 (−1.51, −0.91)	−1.22 (−1.54, −0.84)	−1.27 (−1.51, −1.03)	−1.21 (−1.52, −0.91)
	β_2	0.89 (0.61, 1.15)	0.88 (0.61, 1.15)	0.90 (0.57, 1.23)	0.83 (0.35, 1.29)	0.86 (0.53, 1.18)	0.84 (0.30, 1.38)	0.88 (0.48, 1.25)	0.82 (0.33, 1.30)	0.83 (0.27, 1.38)
	β_3	−1.20 (−1.67, −0.74)	−1.18 (−1.64, −0.71)	−1.26 (−1.84, −0.70)	−1.25 (−1.83, −0.68)	−1.21 (−1.85, −0.59)	−1.34 (−2.03, −0.64)	−1.29 (−2.00, −0.60)	−1.29 (−2.00, −0.57)	−1.38 (−2.18, −0.60)
	τ^2	11.56 (8.99, 15.36)	4.98 (1.24, 11.25)	10.51 (7.75, 14.37)	10.15 (7.55, 13.75)	10.30 (7.67, 13.93)	9.52 (6.67, 13.44)	9.27 (6.92, 13.59)	9.85 (7.16, 13.58)	9.10 (6.19, 13.07)
	$3/\phi_0$	–	29.50 (12.04, 192.98)	20.33 (10.26, 284.90)	21.51 (10.28, 607.29)	22.44 (10.29, 928.79)	20.30 (10.29, 332.59)	21.23 (10.29, 417.25)	21.02 (10.29, 572.52)	20.96 (10.27, 466.56)
	$3/\phi_1$	–	–	24.52 (10.32, 1612.90)	–	–	25.30 (10.32, 1363.64)	32.31 (10.36, 2439.02)	–	25.08 (10.31, 594.06)
	$3/\phi_2$	–	–	–	22.33 (10.31, 699.30)	–	21.35 (10.29, 518.13)	–	22.06 (10.32, 584.80)	22.29 (10.32, 390.12)
	$3/\phi_3$	–	–	–	–	19.52 (10.25, 276.50)	–	20.07 (10.25, 323.62)	19.59 (10.26, 386.25)	19.76 (10.27, 250.63)
	σ_0^2	–	5.98 (1.31, 11.37)	0.36 (0.17, 1.02)	0.36 (0.17, 0.90)	0.36 (0.17, 1.01)	0.36 (0.17, 1.01)	0.36 (0.17, 0.96)	0.37 (0.17, 1.05)	0.36 (0.17, 0.92)
	σ_1^2	–	–	0.36 (0.19, 0.69)	–	–	0.39 (0.21, 0.80)	0.39 (0.20, 0.80)	–	0.40 (0.21, 0.79)
	σ_2^2	–	–	–	0.65 (0.30, 1.52)	–	0.70 (0.32, 1.71)	–	0.64 (0.28, 1.52)	0.70 (0.31, 1.78)
	σ_3^2	–	–	–	–	0.99 (0.37, 2.75)	–	1.08 (0.41, 3.23)	0.94 (0.38, 2.77)	1.07 (0.41, 3.10)
Fit statistics	D	608.41	501.53	600.15	595.76	597.53	588.21	590.97	592.34	583.05
	pD	4.95	41.29	28.61	22.85	19.05	41.54	37.34	29.88	47.55
	DIC	613.37	542.82	628.77	618.61	616.59	629.74	628.31	622.22	630.60
	RMSPE	66.89	67.46	67.53	70.81	68.23	72.41	68.44	73.59	75.28

Table 3
Candidate model parameter estimates and 95% credible intervals for GLEES.

		Non-spatial	SVI	SVC-β_1	SVC-β_2	SVC
Parameter C.I. 50% (2.5%, 97.5%)	β_0	8.27 (7.66, 8.86)	8.40 (7.76, 9.14)	8.31 (7.61, 9.03)	7.94 (7.43, 8.51)	8.06 (7.38, 8.74)
	β_1	−0.62 (−0.79, −0.44)	−0.63 (−0.81, −0.45)	−0.61 (−0.89, −0.31)	−0.61 (−0.77, −0.46)	−0.62 (−0.96, −0.34)
	β_2	0.66 (0.37, 0.96)	0.67 (0.37, 0.99)	0.86 (0.49, 1.23)	0.74 (0.14, 1.29)	0.87 (0.33, 1.45)
	τ^2	4.10 (2.75, 6.63)	0.53 (0.19, 2.01)	0.40 (0.10, 1.67)	0.32 (0.10, 1.22)	0.25 (0.08, 0.95)
	$3/\phi_0$	–	44.11 (12.90, 173.16)	22.85 (10.31, 125.63)	20.96 (10.27, 151.98)	20.19 (10.27, 198.41)
	$3/\phi_1$	–	–	33.88 (10.46, 527.24)	–	29.80 (10.37, 1020.50)
	$3/\phi_2$	–	–	–	58.09 (10.56, 552.49)	28.21 (10.39, 643.78)
	σ_0^2	–	3.45 (1.95, 5.88)	0.98 (0.35, 2.66)	0.71 (0.29, 1.77)	0.53 (0.24, 1.33)
	σ_1^2	–	–	0.44 (0.24, 0.88)	–	0.39 (0.21, 0.85)
	σ_2^2	–	–	–	1.11 (0.54, 2.48)	0.78 (0.35, 1.90)
Fit statistics	D	196.30	103.69	90.35	79.59	67.79
	pD	3.90	29.79	27.58	30.25	33.74
	DIC	200.20	133.48	117.93	109.85	101.53
	RMSPE	35.79	34.21	35.63	35.02	35.20

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.

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 β_2 parameter.

For the FEF, GLEES, and NIWOT sites, the *SVI* models produced marginally to moderately lower *D*, *DIC*, and *RMSPE* as compared to the *non-spatial* models (Tables 1, 3, and 4). These results were not surprising given the clear spatial structure seen in the exploratory semivariograms of the *non-spatial* models' residuals. In contrast, the semivariogram of the *non-spatial* model residuals for MEF did not provide strong evidence of any residual spatial correlation (Fig. 3(b)); hence, the *SVI* model produced only slight gains in model fit, along with a decrease in model prediction accuracy, when compared to the *non-spatial* candidate (Table 2). This demonstrated that the addition of a spatial random effect to the MEF model intercept was not only unnecessary, but allowing the intercept to vary spatially caused the model to over-fit the data and reduced predictive performance.

The candidate models' parameter estimates further corroborated the notion that spatial structure was being absorbed by the spatial random effects in the presence of residual spatial autocorrelation (Tables 1, 2, 3, and 4). The ratio $\tau^2/(\tau^2 + \sigma_0^2)$ calculated using estimates from the *SVI* model proved useful for assessing the need for spatial random effects. Lower ratios indicated greater spatial structure in the *non-spatial* models' residuals and, hence, need to include the spatially-varying intercept. This ratio was 0.15, 0.13, and 0.29 for FEF, GLEES, and NIWOT, respectively. Compared to MEF's 0.45 ratio, the *non-spatial* model residuals for FEF, GLEES, and NIWOT contained more spatial structure.

The effective spatial range is another parameter that describes the strength of a spatial component. A long spatial range suggests stronger and farther reaching spatial dependence. MEF's *SVI* spatial range of 29.5 m was much shorter than the *SVI* ranges for FEF, GLEES and NIWOT (Tables 1, 2, 3, and 4). The long spatial ranges and low nugget to total variance ratios suggested the spatial random effects were capturing substantial residual spatial structure and the *non-spatial* models were not appropriate for the FEF, GLEES, and NIWOT datasets. That is, inference drawn from the *non-spatial* model at FEF, GLEES, and NIWOT—particularly uncertainty for parameter point estimates and predictions—should be interpreted with caution due to the correlated residuals and hence violation of the independent and identically distributed residuals assumption. In contrast, the variance ratio and effective spatial range values suggested that there was little spatial structure in the MEF *non-spatial* model residuals—indicating that spatial random effects were not warranted.

Table 4
Candidate model parameter estimates and 95% credible intervals for NIWOT.

		Non-spatial	SVI	SVC-β_1	SVC-β_2	SVC
Parameter C.I. 50% (2.5%, 97.5%)	β_0	12.34 (11.83, 12.85)	12.35 (11.38, 13.18)	12.35 (11.42, 13.01)	12.19 (11.38, 12.85)	12.21 (11.47, 13.06)
	β_1	−0.82 (−0.95, −0.68)	−0.77 (−0.93, −0.60)	−0.84 (−1.17, −0.17)	−0.77 (−0.93, −0.59)	−0.84 (−1.22, −0.29)
	β_2	0.21 (−0.04, 0.49)	0.21 (−0.05, 0.48)	0.31 (−0.01, 0.62)	0.23 (−0.52, 0.78)	0.34 (−0.17, 0.82)
	τ^2	4.01 (2.82, 5.98)	1.29 (0.36, 3.74)	1.09 (0.25, 2.76)	0.99 (0.21, 3.14)	0.66 (0.20, 2.10)
	$3/\phi_0$	–	116.58 (13.64, 1444.64)	60.19 (10.52, 1578.95)	137.24 (11.40, 1167.32)	42.98 (10.43, 1401.87)
	$3/\phi_1$	–	–	50.58 (10.40, 2097.90)	–	77.34 (10.46, 2419.35)
	$3/\phi_2$	–	–	–	45.68 (10.45, 2343.75)	22.29 (10.30, 669.64)
	σ_0^2	–	2.66 (0.78, 5.07)	1.19 (0.34, 3.21)	1.74 (0.41, 3.87)	0.87 (0.30, 2.51)
	σ_1^2	–	–	0.37 (0.20, 1.03)	–	0.38 (0.20, 1.00)
	σ_2^2	–	–	–	0.74 (0.32, 2.03)	0.78 (0.33, 1.90)
Fit statistics	D	262.40	189.65	177.34	172.77	150.60
	pD	3.93	27.65	34.78	28.97	41.52
	DIC	266.32	217.29	212.12	201.74	192.12
	RMSPE	50.05	48.27	49.31	46.31	49.40

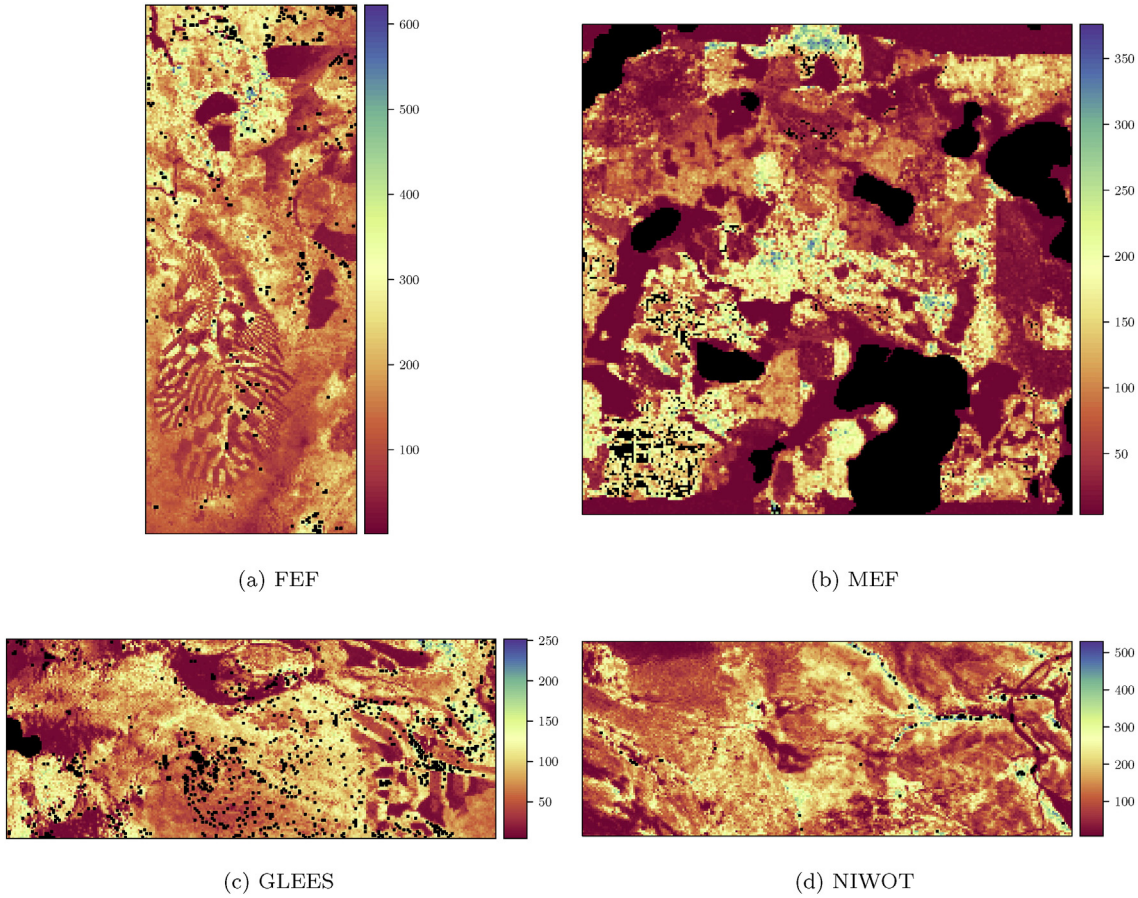


Fig. 4. Maps showing predicted AGB means (Mg/ha) for the full LiDAR extent at FEF, MEF, GLEES, and NIWOT. Black squares indicate excluded prediction grid cells (see Section 2.6). Supplementary data associated with this article can be found, in the online version, at doi:(Please insert doi here).

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*, *DIC*, 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 $\beta_k(\boldsymbol{s}) = \beta_k + w_k(\boldsymbol{s})$ can provide insights into the space-varying relationship between the *k*-th covariate and AGB. For example, the $\beta_1(\boldsymbol{s})$ 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., $\beta_1(\boldsymbol{s})$ was farther from zero in the north (Fig. 6b). After examining the LiDAR derived digital terrain model, it could be seen that $\beta_1(\boldsymbol{s})$ was related to elevation (Fig. 7). This suggests that the relationship between AGB and the LiDAR covariate associated with $\beta_1(\boldsymbol{s})$ 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 β_1 to vary spatially.

Compared to $\beta_1(\boldsymbol{s})$, random effects associated with $\beta_0(\boldsymbol{s})$ and $\beta_2(\boldsymbol{s})$ 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 $\beta(\boldsymbol{s})'s$ through a smoothing filter, the more subtle spatial adjustments to β_0 and β_2 could be seen. Fig. 8a and b depicts $\beta_0(\boldsymbol{s})$ and $\beta_2(\boldsymbol{s})$ 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 $\beta_0(\boldsymbol{s})$ and $\beta_2(\boldsymbol{s})$, confirming that the random effects are capturing local or micro-scale variability. The global estimates of β_0 and β_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

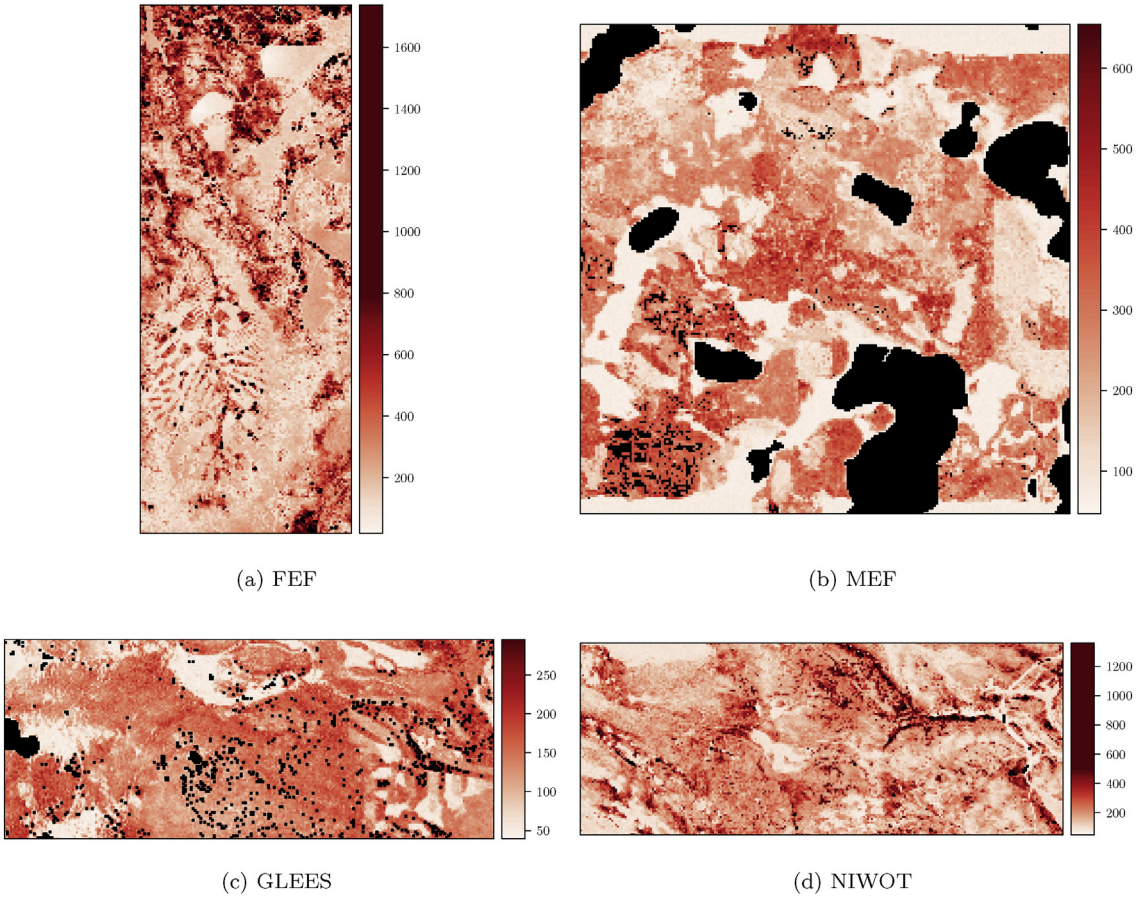


Fig. 5. Maps showing predicted AGB 95% credible interval width (Mg/ha) at FEF, MEF, GLEES, and NIWOT. Black squares indicate excluded prediction grid cells (see Section 2.6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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).

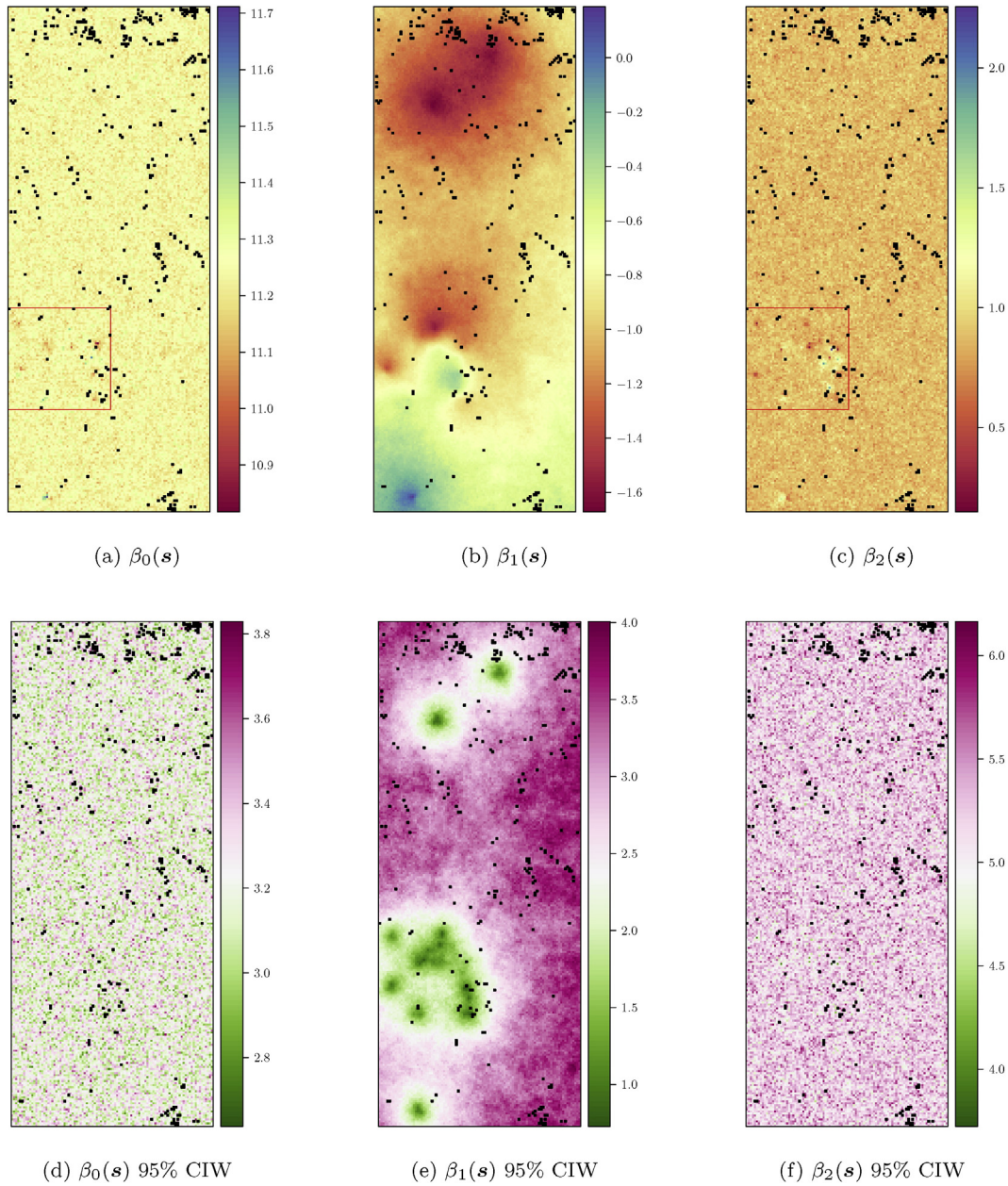
4.4. NIWOT results

The *SVC*-β_2 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*-β_2), but this did not translate into improved prediction accuracy.

NIWOT's *non-spatial* model β_2 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 β_2 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 $\beta_2(\boldsymbol{s})$ than others (Fig. 10b). This result provided evidence that in some locations, the LiDAR covariate associated with β_2 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



(a) $\beta_0(\boldsymbol{s})$ (b) $\beta_1(\boldsymbol{s})$ (c) $\beta_2(\boldsymbol{s})$

(d) $\beta_0(\boldsymbol{s})$ 95% CIW (e) $\beta_1(\boldsymbol{s})$ 95% CIW (f) $\beta_2(\boldsymbol{s})$ 95% CIW

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). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

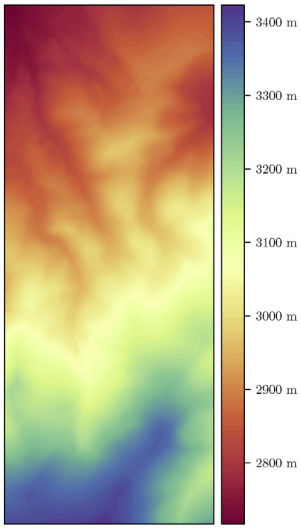


Fig. 7. Digital terrain model for FEF.

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

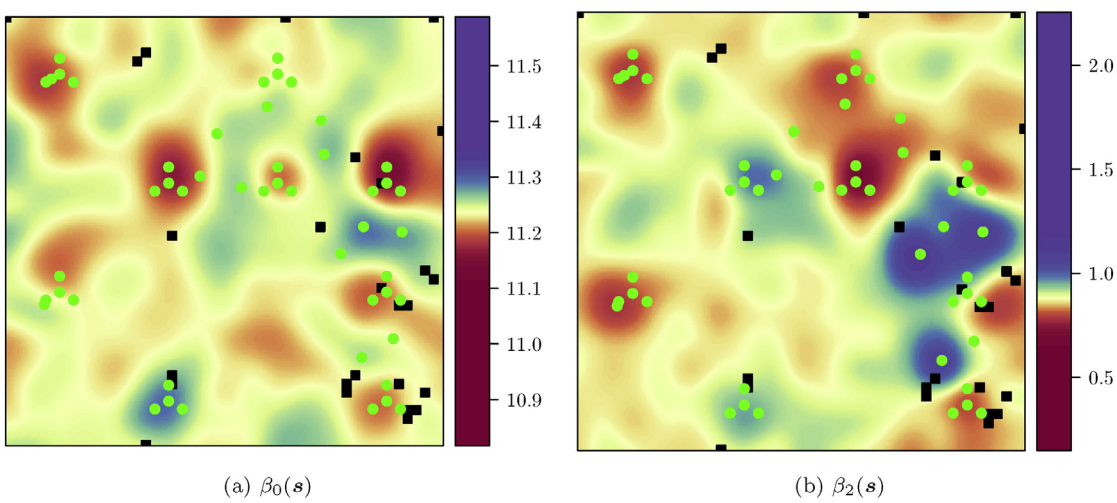


Fig. 8. Maps showing the spatially varying coefficients zoomed into the extent outlined in red in Fig. 6a and 6c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green.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

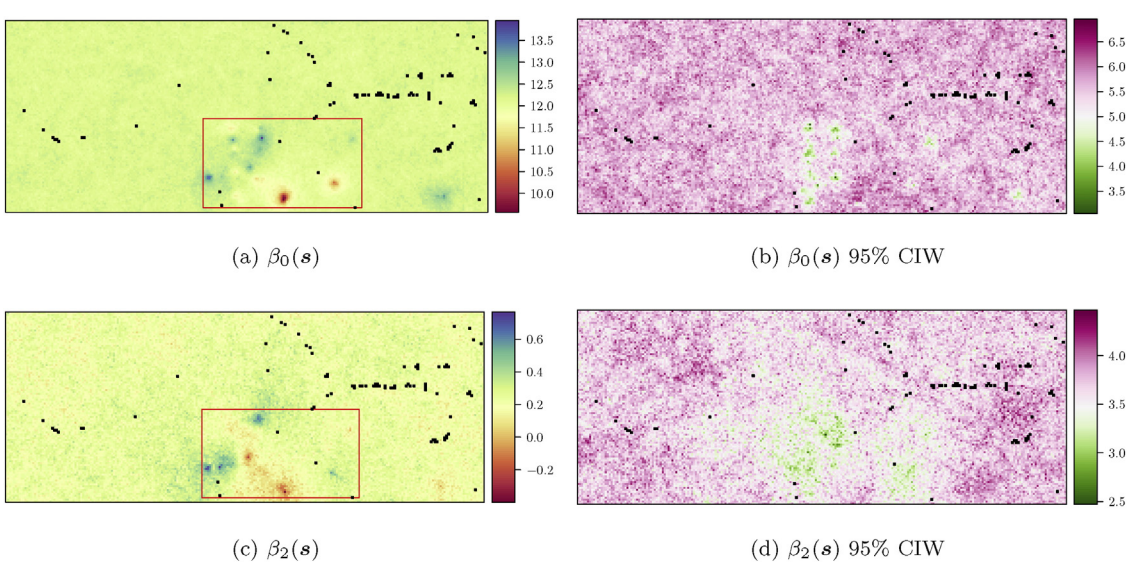
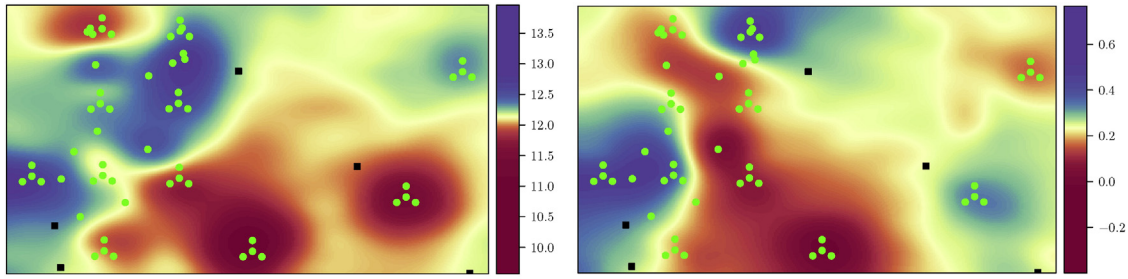


Fig. 9. Maps showing the spatially varying coefficients and associated 95% credible interval width (CIW) for the *SVC*-β2 model at NIWOT. Black squares indicate excluded prediction grid cells (see Section 2.6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) $\beta_0(\boldsymbol{s})$ (b) $\beta_2(\boldsymbol{s})$

Fig. 10. Maps showing the spatially varying coefficients zoomed into the extent outlined in red in Fig. 9a and 9c. 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). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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, should be tested.

Bayesian space-varying coefficient models provide statically valid inference about non-stationary relationships between the covariates and the outcome variable. Results from the NIWOT analysis showed this phenomenon when β_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 NNH08AH971, and the USDA Forest Service. Special thanks to the anonymous reviewers.

Appendix A

A1. Principal components analysis procedure

Let $\boldsymbol{A}$ be the sub-plot and $\boldsymbol{B}$ 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 $\tilde{\boldsymbol{A}}$ be the standardized scores for $\boldsymbol{A}$ calculated as follows,

$$\tilde{\boldsymbol{A}}_{i,j} = \frac{\boldsymbol{A}_{i,j} - \boldsymbol{\mu}_j}{\boldsymbol{\sigma}_j}, \tag{A1}$$

here $\boldsymbol{\mu}_j$ and $\boldsymbol{\sigma}_j$ are the mean and standard deviation of the j-th column of $\boldsymbol{A}$. A correlation matrix $\boldsymbol{\rho}$ is constructed for $\boldsymbol{A}$ by,

$$\boldsymbol{\rho} = \frac{\tilde{\boldsymbol{A}}'\tilde{\boldsymbol{A}}}{n-1}. \tag{A2}$$

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

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

$$\tilde{\boldsymbol{B}}_{i,j} = \frac{\boldsymbol{B}_{i,j} - \boldsymbol{\mu}_j}{\boldsymbol{\sigma}_j}. \tag{A3}$$

Note that $\boldsymbol{\mu}_j$ and $\boldsymbol{\sigma}_j$ are the mean are standard deviation calculated from $\boldsymbol{A}$. Then $\boldsymbol{B}^* = \tilde{\boldsymbol{B}}\boldsymbol{P}$ is then the gridded percentile height data projected into the same vector space as $\boldsymbol{A}^*$.

In practice, a subset of columns from $\boldsymbol{A}^*$ (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 $\boldsymbol{A}^*$ 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 $\boldsymbol{A}^*$ whose standardized eigenvalues total was >0.8. To meet this criterion the two leftmost columns of $\boldsymbol{A}^*$ 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 $\boldsymbol{B}^*$ for each site to use in subsequent grid prediction and mapping.

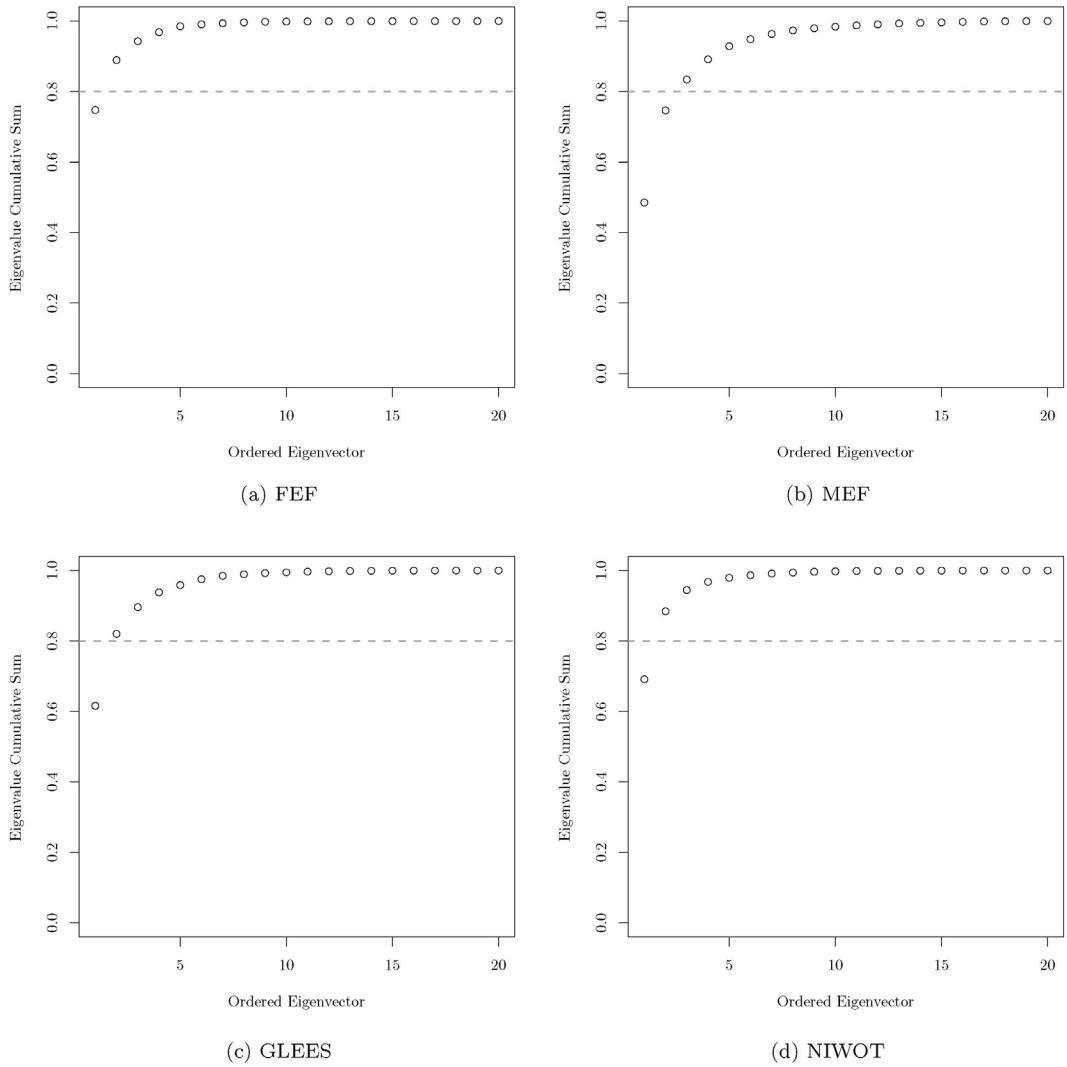


Fig. A1. The cumulative sums of the standardized eigenvalues (in decreasing order) for FEF, MEF, GLEES, and NIWOT.

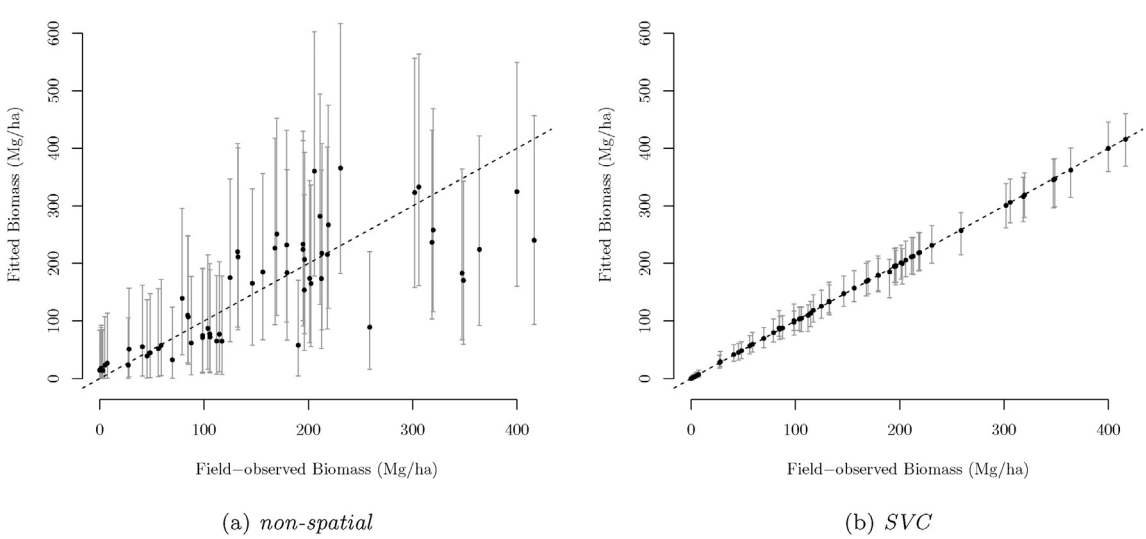


Fig. A2. Plots showing fitted versus field-observed AGB values for FEF's *non-spatial* and *SVC* models. Vertical Gray error bars depict 95% credible intervals for fitted values.

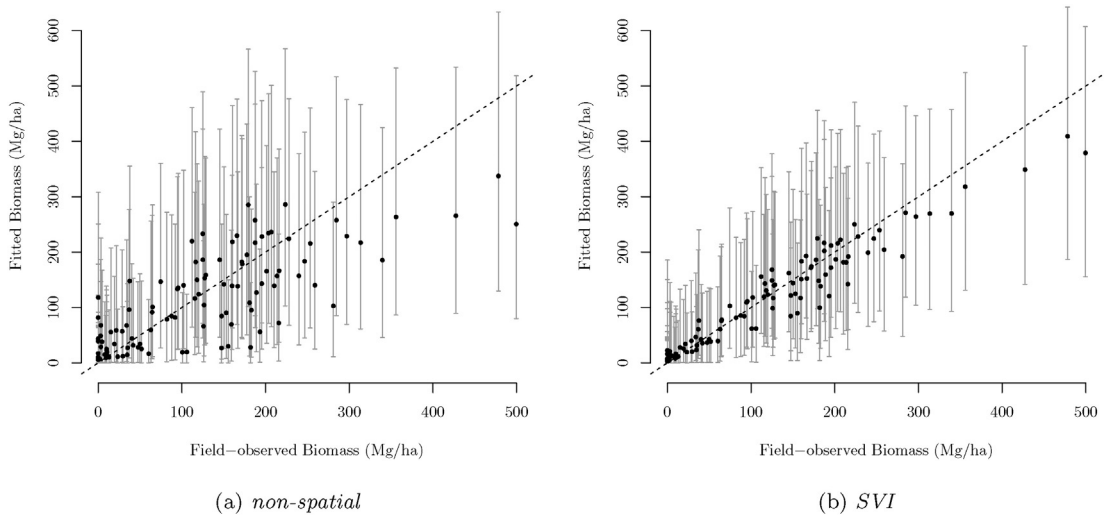


Fig. A3. Plots showing fitted versus field-observed AGB values for MEF's *non-spatial* and *SVI* models. Vertical gray error bars depict 95% credible intervals for fitted values.

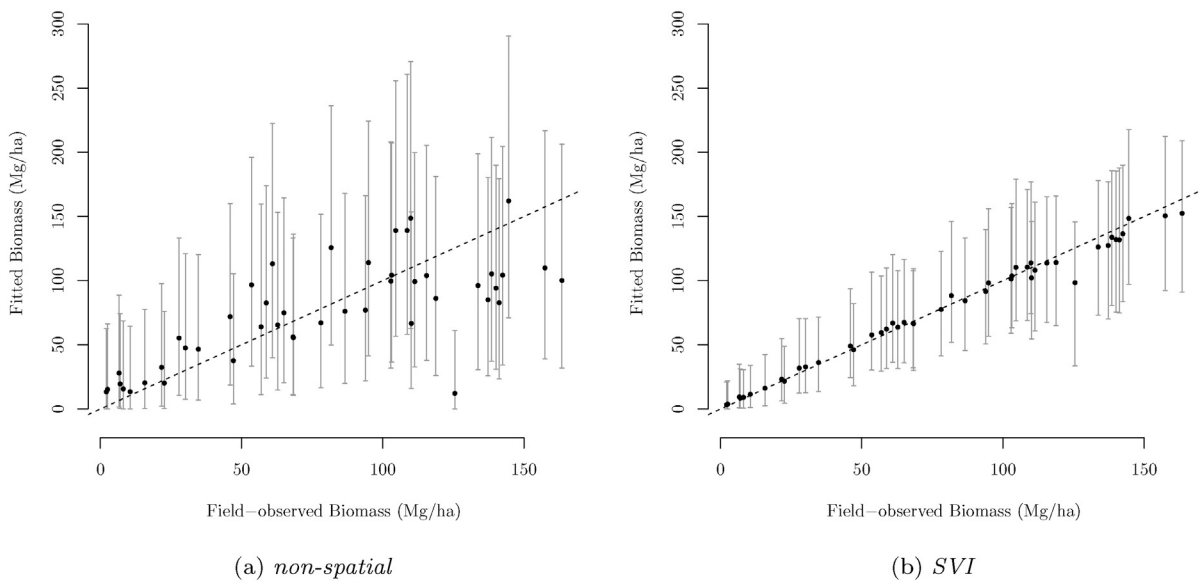


Fig. A4. Plots showing fitted versus field-observed AGB values for GLEES's *non-spatial* and *SVI* models. Vertical gray error bars depict 95% credible intervals for fitted values.

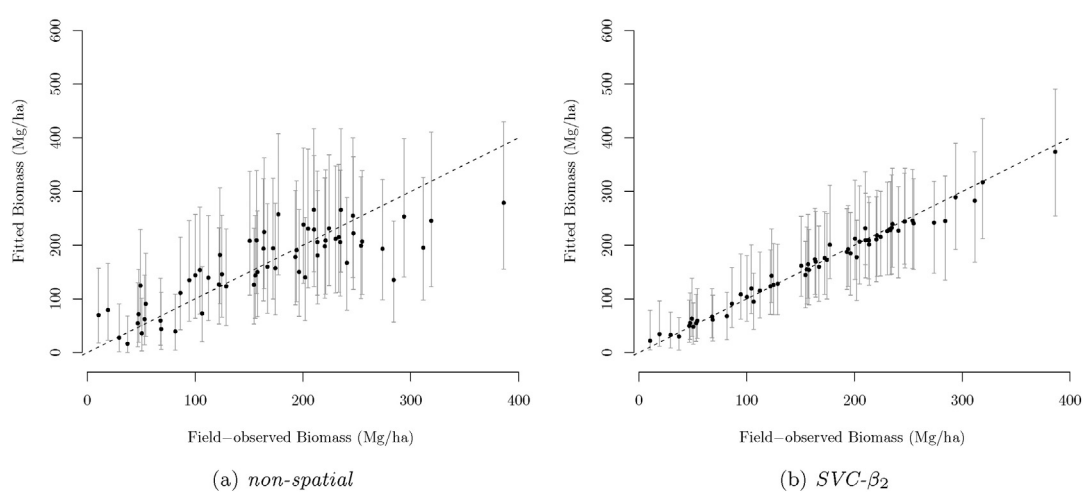


Fig. A5. Plots showing fitted versus field-observed AGB values for NIWOT's *non-spatial* and *SVC-*β_2 models. Vertical gray error bars depict 95% credible intervals for fitted values.

References

Anderson, J., Plourde, L., Martin, M., Braswell, B., Smith, M. L., Dubayah, R., et al. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, *112*, 1856–1870.

Babcock, C., Matney, J., Finley, A.O., Weiskittel, A., & Cook, B.D. (2013). Multivariate spatial regression models for predicting individual tree structure variables using lidar data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*, 6–14.

Banerjee, S., Carlin, C., & Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC.

Banerjee, S., Finley, A., Waldmann, P., & Ericsson, T. (2011). Hierarchical spatial process models for multiple traits in large genetic trials. *Journal of the American Statistical Association*, *105*, 506–521.

Banerjee, S., Gelfand, A., Finley, A., & Huiyan, S. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *70*, 825–848.

Bechtold, W., & Patterson, P. (2005). The enhanced forest inventory and analysis program—National sampling design and estimation procedures. *General technical report SRS*. USDA Forest Service, Southern Research Station.

Bradford, J., Birdsey, R., Joyce, L., & Ryan, M. (2008). Tree age, disturbance history, and carbon stocks and fluxes in subalpine rocky mountain forests. *Global Change Biology*, *14*, 2882–2897.

Bradford, J., Weishampel, P., Smith, M. L., Kolka, R., Birdsey, R., Ollinger, S., et al. (2010). Carbon pools and fluxes in small temperate forest landscapes: Variability and implications for sampling design. *Forest Ecology and Management*, *259*, 1245–1254.

Carlin, B., & Louis, T. (2000). *Bayes and empirical Bayes methods for data analysis* Chapman & Hall/CRC Texts in Statistical Science (2nd ed.). Taylor & Francis.

Chatterjee, S., & Hadi, A. (2006). *Regression analysis by example* Wiley series in probability and statistics (4th ed.). John Wiley & Sons.

Chen, G., Zhao, K., McDermid, G. J., & Hay, G. J. (2012). The influence of sampling density on geographically weighted regression: A case study using forest canopy height and optical data. *International Journal of Remote Sensing*, *33*, 2909–2924.

Chilès, J., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty. Wiley series in probability and statistics*John Wiley & Sons.

CMS (2010). NASA carbon monitoring system. http://carbon.nasa.gov (Accessed: 3-18-2014)

Cole, J. A., Johnson, K. D., Birdsey, R. A., Pan, Y., Wayson, C. A., McCullough, K., Hoover, C. M., Hollinger, D. Y., Bradford, J. B., Ryan, M. G., Kolka, R. K., Wieshampel, P., Clark, K. L., Skowronski, N. S., Hom, J., Ollinger, S. V., McNulty, S. G., & Gavazzi, M. J. (2013). *Database for landscape-scale carbon monitoring sites. Gen. Tech. Rep. NRS-119*. (pp. 12) Newtown square, PA: U.S.: Department of Agriculture, Forest Service, Northern Research Station.

Cressie, N. (1993). *Statistics for spatial data. Wiley series in probability and mathematical statistics: Applied probability and statistics*John Wiley & Sons.

Finley, A. (2010). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, *2*, 143–154.

Finley, A., Banerjee, S., & MacFarlane, D. (2011). A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. *Journal of the American Statistical Association*, *106*, 31–48.

Finley, A., Banerjee, S., & McRoberts, R. (2009). Hierarchical spatial models for predicting tree species assemblages across large domains. *Annals of Applied Statistics*, *3*, 1052–1079.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.

Gelfand, A., Schmidt, A., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, *13*, 263–312.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2013). *Bayesian data analysis* Chapman & Hall/CRC texts in statistical science (3rd ed.). Chapman & Hall/CRC.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.

Gonzalez, P., Asner, G., Battles, J., Lefsky, M., Waring, K., & Palace, M. (2010). Forest carbon densities and uncertainties from lidar, quickbird, and field measurements in California. *Remote Sensing of Environment*, *114*, 1561–1575.

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P., Gómez-Dans, J., Kuze, A., et al. (2012). Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sensing of Environment*, *121*, 236–251.

Guanter, L., Rossini, M., Colombo, R., Meroni, M., Frankenberg, C., Lee, J. E., et al. (2013). Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space. *Remote Sensing of Environment*, *133*, 52–61.

Harville, D. (1997). *Matrix Algebra from a statistician's perspective*. Springer.

Hoeting, J. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, *19*, 574–577.

Huang, B., Song, H., Cui, H., Peng, J., & Xu, Z. (2014). Spatial and spectral image fusion using space matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, *52*, 1693–1704.

Hudak, A. T., Lefsky, M. A., Cohen, W. B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM + data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, *82*, 397–416.

Jarzyna, M. A., Finley, A. O., Porter, W. F., Maurer, B. A., Beier, C. M., & Zuckerberg, B. (2014). Accounting for the space-varying nature of the relationships between temporal community turnover and the environment. *Ecography*, *37*, 1073–1083.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis, Vol. 6*, Pearson Prentice Hall.

Lim, K., & Treitz, P. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research*, *19*, 558–570.

Lucas, R., Cronin, N., Lee, A., Moghaddam, A., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. *Remote Sensing of Environment*, *100*, 407–425.

Muss, J., Mladenoff, D., & Townsend, P. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, *115*, 824–835.

Perrin, E. (1904). On some dangers of extrapolation. *Biometrika*, *3*, 99–103.

Popescu, S., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment*, *115*, 2786–2797.

Richards, J. A. (2013). *Remote sensing digital image analysis, Vol. 5*, Springer.

Salas, C., Ene, L., Gregoire, T. G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, *114*, 1277–1285.

Sherrill, K., Lefsky, M., Bradford, J., & Ryan, M. (2008). Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central Rockies. *Canadian Journal of Forest Research*, *38*, 2081–2096.

Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, *64*, 583–639.

Tonolli, S., Dalponte, M., Neteler, M., Rodeghiero, M., Vescovo, L., & Gianelle, D. (2011). Fusion of airborne lidar and satellite multispectral data for the estimation of timber volume in the Southern Alps. *Remote Sensing of Environment*, *115*, 2486–2498.

UN-REDD (2009). The UN-REDD programme. http://www.un-redd.org/ (Accessed: 3-18-2014)

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications* (3rd ed.). Springer.

Wang, C., Yu, W., Wang, R., Deng, Y., & Zhao, F. (2014). Comparison of nonnegative eigenvalue decompositions with and without reflection symmetry assumptions. *IEEE Transactions on Geoscience and Remote Sensing*, *52*, 2278–2287.

Wheeler, D., & Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, *9*, 145–166.

Wheeler, D., & Waller, L. (2009). Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, *11*, 1–22.