International Journal of Applied Earth Observation and Geoinformation 22 (2013) 147–160

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Hierarchical Bayesian spatial models for predicting multiple forest variables using waveform LiDAR, hyperspectral imagery, and large inventory datasets

Andrew O. Finley[a,*], Sudipto Banerjee[b], Bruce D. Cook[c], John B. Bradford[d]

[a] Departments of Forestry and Geography, Michigan State University, East Lansing, MI, USA
[b] Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA
[c] NASA Goddard Space Flight Center, Biospheric Sciences Laboratory, Greenbelt, MD, USA
[d] U.S. Geological Survey, Southwest Biological Science Center, Flagstaff, AZ, USA

ARTICLE INFO

Article history:
Received 12 July 2011
Accepted 11 April 2012

Keywords:
LiDAR
Hyperspectral
Bayesian hierarchical spatial models
Gaussian Predictive process
Forestry

ABSTRACT

In this paper we detail a multivariate spatial regression model that couples LiDAR, hyperspectral and forest inventory data to predict forest outcome variables at a high spatial resolution. The proposed model is used to analyze forest inventory data collected on the US Forest Service Penobscot Experimental Forest (PEF), ME, USA. In addition to helping meet the regression model's assumptions, results from the PEF analysis suggest that the addition of multivariate spatial random effects improves model fit and predictive ability, compared with two commonly applied modeling approaches. This improvement results from explicitly modeling the covariation among forest outcome variables and spatial dependence among observations through the random effects. Direct application of such multivariate models to even moderately large datasets is often computationally infeasible because of cubic order matrix algorithms involved in estimation. We apply a spatial dimension reduction technique to help overcome this computational hurdle without sacrificing richness in modeling.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Recent advances in remote sensing, specifically Light Detection and Ranging (LiDAR) and hyperspectral sensors, provide detailed data at unprecedented scales. At broad spatial scales, large amounts of LiDAR and hyperspectral data for temperate North American forests are being collected, or will soon be collected by US Forest Service, NASA's Laser Vegetation Imaging Sensor (LVIS; http://lvis.gsfc.nasa.gov), National Ecological Observatory Network (NEON) Airborne Observation Platform (Kamper et al., 2010), and upcoming NASA missions. These high-dimensional data contain information about forest structure and reflectance that can be related to forest variables of interest through suitable modeling frameworks. Anderson et al. (2008) and Tonolli et al. (2011) offer a thorough literature review of recent studies that couple these remotely sensed data with georeferenced forest inventory through parametric or non-parametric regression techniques. The inference garnered through these analyses often supports decisions with important economic and ecological implications; therefore, it is critical to correctly estimate uncertainty.

Commonly, the regression models used to explore relationships between covariates, including remotely sensed variables, and the forest outcome variables of interest do not explicitly accommodate residual spatial dependence. These non-spatial regression models are adequate in the absence of extraneous structured variation, beyond what is explained by the covariates. However, when observations are spatially indexed, we might expect similar outcomes in proximate locations, possibly resulting from common environmental conditions or disturbance regimes. Ignoring this spatial dependence can lead to incorrect inference about model parameters and erroneous predictions. Hoeting (2009) offers a nice discussion on consequences of not meeting the assumption of uncorrelated model residuals. A common solution to accommodating spatial dependence among the residuals is to add a spatially varying model intercept via normally distributed spatial random effects that have a mean of zero and covariance that captures spatial association through a function of distance and perhaps direction between observations. Beyond helping to ensure the statistical validity of the model, the addition of a spatial random effects to the intercept allows for partitioning of residual uncertainty into a spatial and non-spatial component which can improve model fit and predictive performance.

Theoretical and applied studies have shown that spatial associations are captured most effectively using models that build dependencies in different stages. Hierarchical models are especially useful for acknowledging multiple sources of uncertainty and dependence, where they can estimate richer models with less stringent assumptions than traditional modeling paradigms. Flexible

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

0303-2434/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jag.2012.04.007

specification of these models follows the Bayesian paradigm of statistical inference (see, e.g., Carlin and Louis, 2000; Gelman et al., 2004), where analysis uses samples, obtained using Markov chain Monte Carlo (MCMC) methods, from the posterior distributions of model parameters.

Computational advances with regard to MCMC have contributed enormously to the popularity of hierarchical models in a wide array of disciplines (e.g., Gilks et al., 1996; Robert and Casella, 2005), and spatial modeling is no exception (see, e.g., Banerjee et al., 2004). These hierarchical spatial process models, implemented through MCMC methods, offer increased flexibility to fit models that would be infeasible with classical methods. However, with this increased flexibility comes substantial computational demands. Estimating the proposed models involves expensive matrix decompositions whose computational complexity increases in cubic order with the number of spatial locations, rendering such models infeasible for large spatial datasets.

In this paper we detail a spatial regression model that incorporates information from LiDAR and hyperspectral data, often referred to as *fusion* in the remote sensing literature, along with information inherent in the observed forest inventory data to predict multiple forest outcome variables at a high spatial resolution. In particular, we define and assess the utility of: (i) LiDAR and hyperspectral data variable selection using a dimension reduction technique; (ii) multivariate spatial regression with spatial random effects that accommodate dependence of outcomes variables both within and between inventory plot locations, and (iii) a technique called the *predictive process* that reduces the dimensionality of the spatial random effects to ease the computational burden associated with parameter estimation.

The remainder of the paper evolves as follows. The study area and dataset used to illustrate the proposed models are described in Section 2.1. The proposed regression models are detailed in Section 2.2 along with the associated dimension reduction approach and model validation methods. A univariate spatial regression model and k-nearest neighbor algorithm used as benchmarks are defined in Section 2.3. Our approach for LiDAR and hyperspectral variable extraction is provided in Section 2.4. Details about model implementation and subsequent analysis results are given in Sections 3 and 4, respectively. Finally, some concluding remarks with an indication of future work are provided in Section 5.

2. Materials and methods

2.1. Study area and data

The study area is the US Forest Service Penobscot Experimental Forest (PEF; http://www.fs.fed.us/ne/durham/4155/penobsco.htm), ME, USA. The dataset comprised LiDAR waveforms collected with the Laser Vegetation Imaging Sensor (LVIS; http://lvis.gsfc.nasa.gov), hyperspectral imagery from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS; http://aviris.jpl.nasa.gov), and several forest variables measured on a set of 451 georeferenced forest inventory plots.

Both the LVIS and the AVIRIS data were acquired during the summer of 2003. The LVIS instrument, an airborne scanning LiDAR with a 1064 nm laser, provided full waveform LiDAR returns for an approximate 25 m diameter footprint. For each waveform, elevations were converted to height above the ground surface and interpolated at 0.3 m intervals. The AVIRIS instrument provides up-welling radiance in 224 contiguous optical bands from 400 to 2500 nm with a spectral resolution of 10 nm. Radiance data were acquired at a spatial resolution of 18 m and converted to reflectance using the ACORN atmospheric correction program (ACORN 5.0; Ollinger et al., 2008).

The PEF forest inventory data are collected every 5 years as part of a long-term silvicultural treatment study (for more details see Sendak et al., 2003). Our study considered the outcome variables live aboveground tree biomass (BIO; Mg ha^{-1}), basal area (BA; m^2 ha^{-1}), density (DEN; trees ha^{-1}), mean tree diameter at breast height (DBH; cm), and total volume growth between 2002 and 2003 (GRO; Mg ha^{-1}). Live, aboveground biomass was estimated using the generalized biometric models of Jenkins et al. (2004). Growth rates were estimated as differences between estimates of live, aboveground biomass, and biomass trajectories were used to interpolate biomass measurements to the 2003 remote sensing data acquisitions. Fig. 1 provides maps of the outcome variables measured on each forest inventory plot.

A total of 12,414 and 13,403 LiDAR waveform and hyperspectral reflectance signals fell within the PEF boundary. To facilitate model development, hyperspectral data were resampled, using first nearest neighbor, to the LIDAR location grid. Then those signals nearest the center of the inventory plot were extracted for subsequent plot-level analysis.

2.2. Multivariate spatial regression

For the subsequent development, we assume at any location $\boldsymbol{s} \in D \subseteq \Re^2$, we have observed m outcome variables, $y_1(\boldsymbol{s}), y_2(\boldsymbol{s}), \ldots, y_m(\boldsymbol{s})$. For each $y_q(\boldsymbol{s})$, we observe p_q spatially referenced predictors, which we collect into a $p_q \times 1$ vector $\boldsymbol{x}_q(\boldsymbol{s})$. Let $\boldsymbol{y}(\boldsymbol{s}) = (y_1(\boldsymbol{s}), y_2(\boldsymbol{s}), \ldots, y_m(\boldsymbol{s}))'$ be an $m \times 1$ vector with $y_q(\boldsymbol{s})$ as the qth element. We envision a model that accounts for inherent association among the outcomes within each location and among observations across locations. We achieve this through the following multivariate regression model

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

Here $X(\boldsymbol{s})'$ is an $m \times p$ block-diagonal matrix, where $p = \sum_{q=1}^{m} p_q$ and the qth diagonal block is $\boldsymbol{x}_q(\boldsymbol{s})'$. Conformably, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p')'$ is a $p \times 1$ vector of regression coefficients, where each $\boldsymbol{\beta}_q$ is a $p_q \times 1$ vector of regression coefficients corresponding to $\boldsymbol{x}_q(\boldsymbol{s})'$. This specifies the mean structure that accounts for large scale variation in the outcome. Spatial variation in the outcome is modeled using an $m \times 1$ vector of spatially indexed random effects $\boldsymbol{w}(\boldsymbol{s}) = (w_1(\boldsymbol{s}), \ldots, w_m(\boldsymbol{s}))'$. Customarily, we assume the unstructured residuals $\epsilon(\boldsymbol{s})$, defined analogous to $\boldsymbol{w}(\boldsymbol{s})$, follow a zero-centered multivariate normal (MVN) distribution with zero mean and an $m \times m$ covariance matrix Ψ.

Spatially structured dependence is introduced in (1) through a multivariate $(m \times 1)$ *Gaussian process* $\boldsymbol{w}(\boldsymbol{s}) \sim GP(\boldsymbol{0}, C(\cdot, \cdot; \boldsymbol{\theta}))$ (see, e.g., Cressie, 1993), where the *cross-covariance function* $C(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\theta})$ is defined to be the $m \times m$ matrix with (i, j)th entry $\mathrm{Cov}\{w_i(\boldsymbol{s}_1), w_j(\boldsymbol{s}_2)\}$. The cross-covariance function is parametric, with $\boldsymbol{\theta}$ denoting the set of parameters that describe association structure between $\boldsymbol{w}(\boldsymbol{s}_1)$ and $\boldsymbol{w}(\boldsymbol{s}_2)$. The cross-covariance function completely determines the joint dispersion structure implied by the spatial process.

Consider any finite collection of locations $\mathcal{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ and let $\boldsymbol{w}$ be the $mn \times 1$ vector obtained by stacking up the $\boldsymbol{w}(\boldsymbol{s}_i)$'s one above the other. Then, the distribution of $\boldsymbol{w}$ is multivariate normal with zero mean and an $nm \times nm$ variance covariance matrix $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})$. This matrix is best described as a partitioned matrix composed of $m \times m$ blocks, where the (i, j)th block is $C(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta})$.

Since $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})$ must be symmetric and positive-definite, it is immediate that the cross-covariance function must satisfy (i) $C(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}) = C(\boldsymbol{s}_j, \boldsymbol{s}_i; \boldsymbol{\theta})'$ for any pair of locations, and (ii) $\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{v}_i' C(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}) \boldsymbol{v}_j > 0$ for all non-zero $m \times 1$ vectors $\boldsymbol{v}_i, \boldsymbol{v}_j$. The first condition ensures that $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})$ will be symmetric, although the cross-covariance function itself need not be. The second

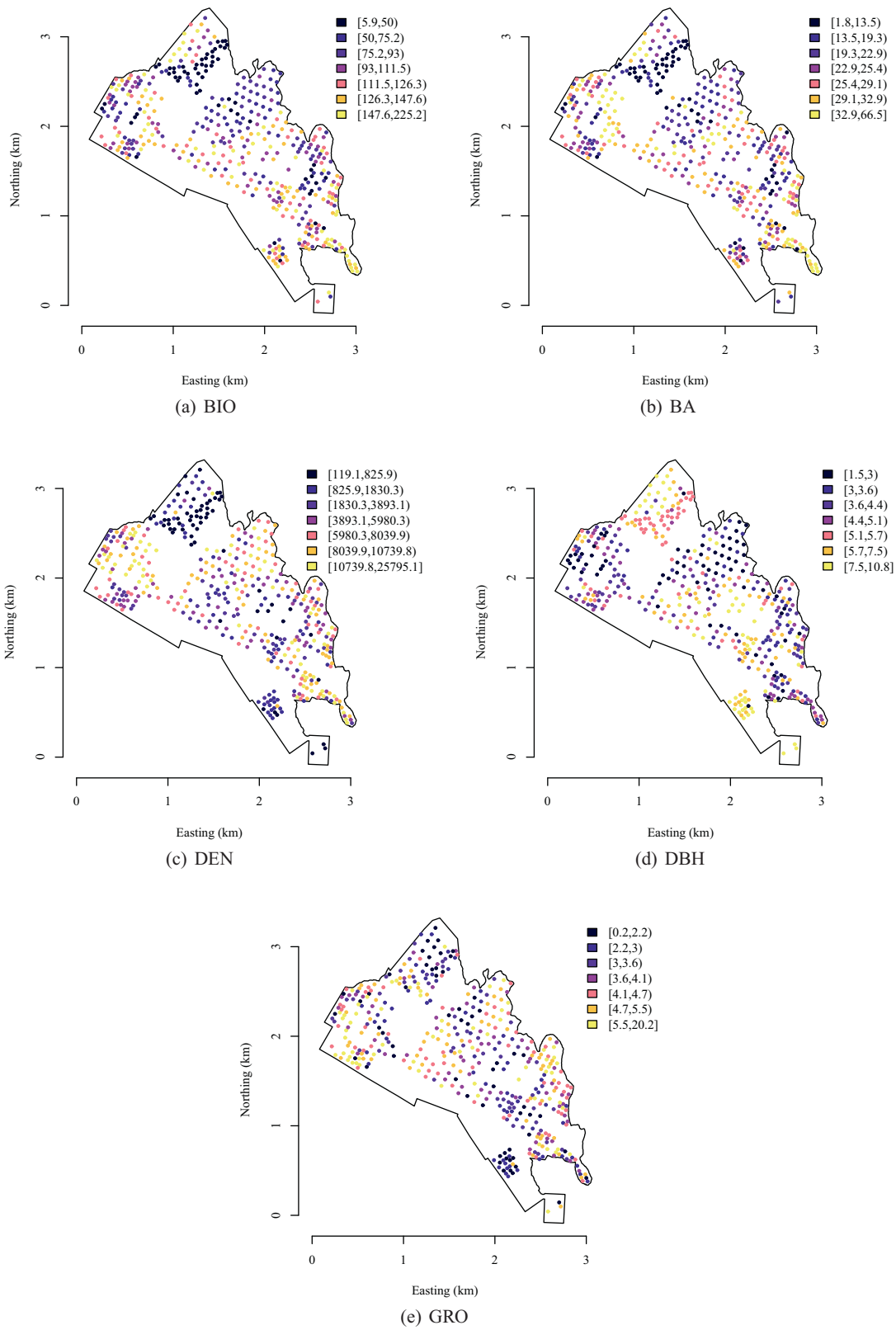

Fig. 1. Forest outcome variables measured at the forest inventory plot locations.

condition ensures the positive-definiteness of $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})$. These conditions must hold for any arbitrary set of distinct locations. The above definition also implies that within a given location $\boldsymbol{s}$, $C(\boldsymbol{s}, \boldsymbol{s}; \boldsymbol{\theta})$ is a symmetric and positive definite matrix.

Let us turn to specifications for cross-covariance functions. One possibility for $C(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\theta})$ is a *separable* specification, $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi})\Sigma$, where $\boldsymbol{\theta} = \{\Sigma, \boldsymbol{\phi}\}$, Σ is an $m \times m$ covariance matrix between the random effects and $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi})$ is a real-valued spatial

correlation function. One important example is the stationary Matérn correlation function

$$\rho(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi}) = \frac{1}{2^{\nu-1}\Gamma(\nu)}(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|\phi)^{\nu}\mathcal{K}_{\nu}(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|\phi); \quad \phi > 0,\ \nu > 0, \tag{2}$$

where $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ is the Euclidean distance between locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, $\boldsymbol{\phi} = \{\phi, \nu\}$ with ϕ controlling the decay in spatial correlation and ν controlling process smoothness, Γ is the usual Gamma function while $\mathcal{K}_{\nu}$ is a modified Bessel function of the second kind with order ν (Cressie, 1993). This implies $\mathrm{Var}(\boldsymbol{w}) = R(\boldsymbol{\phi}) \otimes \Sigma$, where $\otimes$ is the Kronecker product of matrices and $R(\boldsymbol{\phi})$ is the $n \times n$ matrix whose (i, j)th element is $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\phi})$. Note that this separable structure "separates" (or decomposes) the covariance structure into a purely spatial component ($R(\boldsymbol{\phi})$) and a non-spatial component (Σ). The matrix $R(\boldsymbol{\phi})$ captures spatial association, while Σ captures the association among the m outcomes within a location. The separable model is "stationary" in the sense that the association among the outcomes remains the same for every location.

A separable spatial association structure is rather restrictive here, assigning a single set of spatial correlation parameters $\boldsymbol{\phi}$ to be shared by all the outcomes. This can be inappropriate – why should all the outcomes have, for example, the same strength of spatial associations? More general cross-covariance functions, however, are not routine to specify although there is now a fairly rich literature on them. Constructive approaches, dating back to Gelfand et al. (2004), are more comprehensible and computationally feasible. This can be motivated as below.

For the separable model, notice that we can write $C(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\theta}) = A\Theta(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi})A'$ where $\Sigma = AA'$ is a Cholesky factorization, $\Theta(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi})$ is a diagonal matrix with $\rho_q(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\phi}_q)$, $q = 1, \ldots, m$ as its diagonal elements, and $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_m\}$. This incorporates a set of m correlation parameters, or even m different correlation functions, offering an attractive, easily interpretable and flexible approach. This approach resembles the linear model of coregionalization (LMC) as in, e.g., Grzebyk and Wackernagel (1994), Wackernagel (2003), Schmidt and Gelfand (2003), and Gelfand et al. (2004). See, also, Reich and Fuentes (2007) for a Bayesian nonparametric adaptation. We will assume this structure in our subsequent development. Under this setting $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta}) = (I \otimes A)\Sigma_{\Theta}(\boldsymbol{\phi})(I \otimes A')$, where $\Sigma_{\Theta}(\boldsymbol{\phi})$ is an $mn \times mn$ block matrix whose (i, j)th block is $\Theta(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\phi})$, and $\boldsymbol{\theta} = \{A, \boldsymbol{\phi}\}$.

Let $\boldsymbol{y}$ be the $mn \times 1$ vector formed by stacking up the $m \times 1$ $\boldsymbol{y}(\boldsymbol{s}_i)$'s. The posterior distribution $p(\boldsymbol{\beta}, \boldsymbol{w}, \Psi, \boldsymbol{\theta} \mid \boldsymbol{y})$ is proportional to

$$\prod_{q=1}^{m} p(\boldsymbol{\phi}_q) \times IW(AA' \mid a_{AA'}, b_{AA'}) \times MVN(\boldsymbol{\beta} \mid \boldsymbol{\mu}_{\beta}, \Sigma_{\beta}) \times IW(\Psi \mid a_{\Psi}, b_{\Psi}) \times MVN(\boldsymbol{w} \mid \boldsymbol{0}, \Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})) \times \prod_{i=1}^{n} MVN(\boldsymbol{y}(\boldsymbol{s}_i) \mid X(\boldsymbol{s}_i)'\boldsymbol{\beta} + \boldsymbol{w}(\boldsymbol{s}_i), \Psi), \tag{3}$$

where IW is the inverse-Wishart prior distribution with degrees of freedom $a_{AA'}$ and scale matrix $b_{AA'}$. The prior distributions for $\boldsymbol{\phi}$'s are dependent on the chosen spatial correlation function and the extent of the domain.

Choices for modeling A include an inverse-Wishart prior for AA' (Gelfand et al., 2004), as in our current analysis, element-wise modeling with normal and log-normal priors, or using parametric association structures suggested by the design under consideration (see, e.g., Banerjee and Johnson, 2006). Alternatively, instead of a lower-triangular Cholesky factor, one might opt for a spectral square-root (see, e.g., Daniels and Kass, 1999). In previous analyses, we have experimented with each of these approaches and found that posterior estimates of AA' do not vary greatly among the specifications.

In practical implementation, a *marginalized* version of (3) that integrates out $\boldsymbol{w}$ is often preferred because of the smaller parameter space. This uses $MVN(\boldsymbol{y} \mid X\boldsymbol{\beta}, \Sigma_{\boldsymbol{y}}(\boldsymbol{\theta}))$ as the likelihood, where X is now the $nm \times p$ matrix of predictors obtained by stacking the $X(\boldsymbol{s}_i)'$'s, and $\Sigma_{\boldsymbol{y}} = (I_n \otimes A)\Sigma_{\Theta}(\boldsymbol{\phi})(I_n \otimes A') + I_n \otimes \Psi$. The parameters to be updated in the marginalized model are $\Omega = \{\boldsymbol{\beta}, \Psi, A, \boldsymbol{\phi}\}$ whose priors are exactly as in (3).

An efficient MCMC algorithm is obtained by updating $\boldsymbol{\beta}$ from its full conditional distribution $MVN(\boldsymbol{\mu}_{\boldsymbol{\beta}|\cdot}, \Sigma_{\boldsymbol{\beta}|\cdot})$, where

$$\Sigma_{\boldsymbol{\beta}|\cdot} = [\Sigma_{\boldsymbol{\beta}}^{-1} + X'\Sigma_{\boldsymbol{y}}^{-1}X]^{-1} \quad \text{and} \quad \boldsymbol{\mu}_{\boldsymbol{\beta}|\cdot} = \Sigma_{\boldsymbol{\beta}|\cdot}X'\Sigma_{\boldsymbol{y}}^{-1}\boldsymbol{y}. \tag{4}$$

The remaining parameters are updated by using Metropolis steps, possibly using block updates. Typically, random-walk Metropolis steps with (multivariate) normal proposals are adopted; since all parameters with positive support are converted to their logarithms, some Jacobian computation is needed. Upon convergence, the MCMC output generates L samples, say $\{\Omega^{(l)}\}_{l=1}^{L}$, from the parameters' posterior distribution.

In this sampling scheme, the spatial random effects $\boldsymbol{w}$ are not sampled directly. This reduces the parameter space, which results in a more efficient MCMC algorithm. A key advantage of the first stage Gaussian model (as in (1)) is that samples from the posterior distribution of $\boldsymbol{w}$ can be recovered in a posterior predictive fashion. More precisely, we seek to evaluate

$$p(\boldsymbol{w} \mid \boldsymbol{y}) \propto \int p(\boldsymbol{w} \mid \Omega, \boldsymbol{y})p(\Omega \mid \boldsymbol{y})d\Omega. \tag{5}$$

Because the full conditional distribution of $\boldsymbol{w}$ in (5) is again multivariate normal, (5) is easily evaluated using *composition sampling*. To be precise, for each sample $\Omega^{(l)}$, we draw $\boldsymbol{w}^{(l)}$ from $MVN(\boldsymbol{\mu}_{\boldsymbol{w}|\cdot}, \Sigma_{\boldsymbol{w}|\cdot})$, where

$$\Sigma_{\boldsymbol{w}|\cdot} = [(I_n \otimes A^{-1(l)})\Sigma_{\Theta}(\boldsymbol{\phi}^{(l)})^{-1}(I_n \otimes A^{-1(l)\prime}) + I_n \otimes \Psi^{-1(l)}]^{-1}$$

$$\text{and} \quad \boldsymbol{\mu}_{\boldsymbol{w}|\cdot} = \Sigma_{\boldsymbol{w}|\cdot}(I_n \otimes \Psi^{-1(l)})(\boldsymbol{y} - X\boldsymbol{\beta}^{(l)}).$$

These posterior samples from $\boldsymbol{w}$ can be mapped using contours or interpolation routines to produce plots of the spatial process.

For predictions, if $\mathcal{S}_0 = \{\boldsymbol{s}_{0,1}, \boldsymbol{s}_{0,2}, \ldots, \boldsymbol{s}_{0,n_0}\}$ is a collection of n_0 *new* locations, the posterior predictive distribution of $p(\boldsymbol{w}_0 \mid \boldsymbol{y})$, where $\boldsymbol{w}_0 = (w(\boldsymbol{s}_{0,1}), w(\boldsymbol{s}_{0,2}), \ldots, w(\boldsymbol{s}_{0,n_0}))'$, is

$$p(\boldsymbol{w}_0 \mid \boldsymbol{y}) \propto \int p(\boldsymbol{w}_0 \mid \boldsymbol{w}, \Omega, \boldsymbol{y})p(\boldsymbol{w} \mid \Omega, \boldsymbol{y})p(\Omega \mid \boldsymbol{y})d\Omega d\boldsymbol{w}. \tag{6}$$

Given posterior samples, $\{\Omega^{(l)}\}_{l=1}^{L}$, this distribution can again be obtained via composition sampling: we first draw $\boldsymbol{w}^{(l)}$ for each l as described in (5) and then draw $\boldsymbol{w}_0^{(l)}$ from $p(\boldsymbol{w}_0 \mid \boldsymbol{w}^{(l)}, \Omega^{(l)}, \boldsymbol{y})$, where this last distribution is derived as a conditional distribution from a multivariate normal and, hence, is again 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, $p(\boldsymbol{w}_0 \mid \boldsymbol{w}, \Omega, \boldsymbol{y}) = p(\boldsymbol{w}_0 \mid \boldsymbol{w}, \Omega)$, which is a $MVN(\boldsymbol{\mu}_{\boldsymbol{w}_0|\cdot}, \Sigma_{\boldsymbol{w}_0|\cdot})$, where

$$\boldsymbol{\mu}_{\boldsymbol{w}_0|\cdot} = \Sigma_{\boldsymbol{w}_0}(\boldsymbol{\theta})'\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})^{-1}\boldsymbol{w} \quad \text{and}$$

$$\Sigma_{\boldsymbol{w}_0|\cdot} = \Sigma - \Sigma_{\boldsymbol{w}_0}(\boldsymbol{\theta})'\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})^{-1}\Sigma_{\boldsymbol{w}_0}(\boldsymbol{\theta}), \tag{7}$$

$\Sigma_{\boldsymbol{w}_0}(\boldsymbol{\theta})'$ is an $m \times mn$ block matrix whose (j)th block is the $m \times m$ cross-covariance $C(\boldsymbol{s}_0, \boldsymbol{s}_j; \boldsymbol{\theta})$, $j = 1, \ldots, n$.

Finally, given a set of covariates at the new location $\boldsymbol{s}_0$, samples from the posterior predictive distribution of the outcome variable

vector, $\boldsymbol{y}(\boldsymbol{s}_0)^{(l)}$, are drawn from $MVN(X(\boldsymbol{s}_0)'\boldsymbol{\beta}^{(l)} + \boldsymbol{w}(\boldsymbol{s}_0)^{(l)}, \Psi^{(l)})$ for $l = 1, 2, \ldots, L$.

2.2.1. Spatial process models for large datasets

Without further specifications, estimating the parameters in (1) requires linear solvers or Cholesky decompositions of the $mn \times mn$ matrix $\Sigma_{\boldsymbol{w}}(\boldsymbol{\theta})$. The computational complexity is $O(n^3m^3)$ and this work needs to be executed once every MCMC iteration, when we want to estimate $\boldsymbol{\theta}$. With large n and/or m, this is computationally infeasible and we must resort to some approximation or dimension reduction approach. In most practical settings, we encounter large n rather than large m. We now discuss dimension reduction in n.

Modeling large spatial datasets has received much attention in the statistical literature. Vecchia (1988) proposed approximating the likelihood with a product of appropriate conditional distributions to obtain maximum-likelihood estimates. Stein et al. (2004) adapt this to restricted maximum likelihood estimation. Another possibility is to approximate the likelihood using spectral representations of the spatial process (Fuentes, 2007). These likelihood approximations yield a joint distribution, but not a process that facilitates spatial interpolation. Yet another approach considers compactly supported correlation functions (Furrer et al., 2006; Kaufman et al., 2008; Du et al., 2009) that yield sparse correlation structures. More efficient sparse solvers can then be employed for kriging and variance estimation, but the tapered structures may limit modeling flexibility. Also, full likelihood-based inference still requires determinant computations that may be problematic.

Alternatively, one could build models especially geared toward handling of large spatial datasets. These are representations of the spatial process in a lower-dimensional subspace and are often referred to as low-rank or reduced-rank spatial models (Higdon, 2002; Kamman and Wand, 2003; Stein et al., 2007; Stein, 2008; Cressie and Johannesson, 2007; Banerjee et al., 2008; Crainiceaniu et al., 2008). Many of these methods are variants of the so-called "subset of regressors" methods used in Gaussian process regressions for large datasets in machine learning (e.g., Rasmussen and Williams, 2006). The idea here is to consider a smaller set of locations, or "knots", say $\mathcal{S}^* = \{\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_{n*}^*\}$, where the number of knots, n^*, is *fixed* to be much smaller than the number of observed sites, and to express the spatial process realizations over n locations in terms of its realizations over the smaller set of knots. Here, the assumption is there will be an insignificant loss of information about the underlying spatial process given an adequate number of knots that cover the domain.

Recently Banerjee et al. (2008) proposed a class of knot-based spatial process models for large spatial datasets. These models consider a fixed set of $\mathcal{S}^*$, again with $n^* \ll n$, which may or may not form a subset of the entire collection of observed locations in $\mathcal{S}$. Letting $\boldsymbol{w}^* = (\boldsymbol{w}(\boldsymbol{s}_1^*)', \ldots, \boldsymbol{w}(\boldsymbol{s}_n^*)')'$ be the realizations of the *parent process* over $\mathcal{S}^*$, we formulate a *multivariate predictive process* defined as

$$\begin{aligned}\tilde{\boldsymbol{w}}(\boldsymbol{s}) &= E\{w(\boldsymbol{s}) \mid \boldsymbol{w}^*\} = \mathrm{Cov}(\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}^*)\mathrm{Var}^{-1}(\boldsymbol{w}^*)\boldsymbol{w}^* \\ &= \mathcal{C}(\boldsymbol{s};\boldsymbol{\theta})'\Sigma^*(\boldsymbol{\theta})^{-1}\boldsymbol{w}^*, \end{aligned} \tag{8}$$

where $\mathcal{C}(\boldsymbol{s};\boldsymbol{\theta})'$ is an $m \times mn^*$ block matrix with the $m \times m$ matrix $C(\boldsymbol{s}, \boldsymbol{s}_j^*;\boldsymbol{\theta})$ being the jth block, and $\Sigma^*(\boldsymbol{\theta})$ is the $n^*m \times n^*m$ covariance matrix of $\boldsymbol{w}^*$, i.e., with $C(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*;\boldsymbol{\theta})$ as its (i,j)th block. Eq. (8) reveals $\tilde{\boldsymbol{w}}(\boldsymbol{s})$ to be a zero mean $m \times 1$ spatial process with cross-covariance matrix given by $C_{\tilde{\boldsymbol{w}}}(\boldsymbol{s}_i, \boldsymbol{s}_j;\boldsymbol{\theta}) = \mathcal{C}(\boldsymbol{s}_i;\boldsymbol{\theta})'\Sigma^*(\boldsymbol{\theta})^{-1}\mathcal{C}(\boldsymbol{s}_j;\boldsymbol{\theta})$. Again, the appeal behind this formulation is that every spatial process (parent) model produces a corresponding predictive process version. Thus, the predictive process counterpart of the parent model in (1), and subsequently in (3), is derived by replacing $\boldsymbol{w}(\boldsymbol{s})$ with $\tilde{\boldsymbol{w}}(\boldsymbol{s})$. No new parameters are introduced and the cross-covariance function of the predictive process derives directly from that of the parent process.

Finley et al. (2009) identified a systematic bias in the covariance components of multivariate predictive process model, which are corrected by adding a spatially adaptive nugget process to $\tilde{\boldsymbol{w}}(\boldsymbol{s})$. More precisely, we correct this bias by replacing $\tilde{\boldsymbol{w}}(\boldsymbol{s})$ with $\tilde{\boldsymbol{w}}_\epsilon(\boldsymbol{s}) = \tilde{\boldsymbol{w}}(\boldsymbol{s}) + \tilde{\boldsymbol{\epsilon}}(\boldsymbol{s})$, where $\tilde{\boldsymbol{\epsilon}}(\boldsymbol{s}) \overset{iid}{\sim} MVN\{\boldsymbol{0}, C(\boldsymbol{s},\boldsymbol{s};\boldsymbol{\theta}) - \mathcal{C}(\boldsymbol{s};\boldsymbol{\theta})'\Sigma^*(\boldsymbol{\theta})^{-1}\mathcal{C}(\boldsymbol{s};\boldsymbol{\theta})\}$. Details on the algorithms to estimate these bias-corrected multivariate predictive process models can be found in Finley et al. (2009).

Key variables in predictive process modeling are the number and location of knots. The choice of n^* is governed by computational cost and sensitivity to choice. Customarily, the analysis is implemented over different choices of n^* and knot locations. The issue is not dissimilar to a spatial design problem (e.g., Nychka and Saltzman, 1998; Xia and Gelfand, 2006; Diggle and Lophaven, 2006). The standard method is to experiment with different computationally feasible knot intensities and configurations.

2.2.2. Model selection

There are several approaches to assessing model performance and subsequent comparison. Here we will consider the popular *Deviance Information Criterion*, or *DIC*, to rank models in terms of how well they fit the data (Spiegelhalter et al., 2002). This criterion is the sum of the Bayesian deviance (a measure of model fit) and the (effective) number of parameters p_D (a penalty for model complexity). Here, lower *DIC* indicates *better* fit. In addition to *DIC*, we also consider an approach that is based on simulating *independent* replicates for each observed outcome. For example, for the full multivariate spatial model (1), and each location $\boldsymbol{s}_i$ we compute, for each $i = 1, 2, \ldots, n$,

$$\begin{aligned} p(\boldsymbol{y}_{rep}(\boldsymbol{s}_i) \mid \boldsymbol{y}) = \int & N(\boldsymbol{y}_{rep}(\boldsymbol{s}_i) \mid X(\boldsymbol{s}_i)'\boldsymbol{\beta} + \boldsymbol{w}(\boldsymbol{s}_i), \Psi) \\ & \times p(\Omega, \boldsymbol{w}(\boldsymbol{s}_i) \mid \boldsymbol{y}) d\Omega\, d\boldsymbol{w}(\boldsymbol{s}_i) \end{aligned}$$

where the conditional distribution of the replicate $\boldsymbol{y}_{rep}(\boldsymbol{s}_i)$ given the parameters is simply the likelihood component corresponding to $\boldsymbol{y}(\boldsymbol{s}_i)$. Simulation from other candidate models, such as the non-spatial, i.e., $\boldsymbol{w} = \boldsymbol{0}$ or the predictive process models, follow this same approach. Then, letting $\boldsymbol{\mu}_{rep,i}$ and $\boldsymbol{\Sigma}_{rep,i}$ be the posterior predictive mean and variance for each $\boldsymbol{y}_{rep}(\boldsymbol{s}_i)$, we prefer models that excel under a decision-theoretic balanced loss function, penalizing both departure of replicated means from their observed values (lack of fit) and excessive uncertainty in the replicated data. Using a squared error loss function (e.g., Gelfand and Ghosh, 1998), the measures for these two criteria are evaluated as $G = \sum_{i=1}^n \|\boldsymbol{y}(\boldsymbol{s}_i) - \boldsymbol{\mu}_{rep,i}\|^2$, where $\|\cdot\|$ is the standard Euclidean norm, and $P = \sum_{i=1}^n \mathrm{tr}(\boldsymbol{\Sigma}_{rep,i})$. We will use the score $D = G + P$ as a model selection criterion, with lower values of D indicating better models.

To facilitate comparison with similar studies that propose models for coupling remotely sensed and forest inventory data we also consider the coefficient of determination for each outcome variable, $R_q^2 = 1 - \sum_{i=1}^n (y_q(\boldsymbol{s}_i) - \tilde{y}_q(\boldsymbol{s}_i))^2 / \sum_{i=1}^n (y_q(\boldsymbol{s}_i) - \bar{y}_q)^2$ for $q = 1, \ldots, m$. Here, for the qth outcome variable, $\tilde{y}_q(\boldsymbol{s}_i)$ is the ith *fitted* value and $\bar{y}_q$ is the mean over the observed values.

Finally, because forest variable prediction is our overarching goal, we assess model predictive performance using a split-set validation approach. Here, 90% of the observations are used to estimate candidate models' parameters and the remaining 10% are used for subsequent prediction. Models are ranked based on their holdout set root mean squared prediction error (RMSPE), where again lower values of RMSPE indicate greater predictive performance.

2.3. Benchmark models

The simpler alternative to the multivariate model is to specify m univariate spatial regression models that do not explicitly model the covariance among their respective spatial random effects. Therefore, to assess the usefulness of (1), we consider the univariate model counterpart for each outcome using the univariate bias-corrected predictive process approximation. Model specification and prediction follows from Section 2.2 by simply setting $m = 1$ (see, e.g., Banerjee et al., 2008; Finley et al., 2009).

We also consider a multivariate but non-spatial version of (1) where $\boldsymbol{w} = \boldsymbol{0}$ and covariance among the outcomes' residuals is still estimated via Ψ. Additionally, we fit the univariate non-spatial regression model to each outcome.

In addition to the regression models, we also consider a popular non-parametric k-nearest neighbor (k-NN) algorithm for prediction (see, e.g., Tomppo et al., 2008; McRoberts, 2011). To predict a vector of outcome variables at the new location, $\boldsymbol{s}_0$, the most basic k-NN algorithm averages over the corresponding outcomes associated with the k nearest observations locations in $\mathcal{S}$, where *nearest* is measured as the Euclidean distance between the $\boldsymbol{x}(\boldsymbol{s}_0)$ and $\boldsymbol{x}(\boldsymbol{s}_i)$ for $i = 1, \ldots, n$. In practice, the value of k that minimizes some objective function is chosen. Using the split-set validation described in Section 2.2.2, we consider values of k from 1 to 20 and select the one that provides the smallest holdout set RMSPE for the majority of the outcome variables.

2.4. Selection of LiDAR and hyperspectral covariates

The high-dimensional waveform LiDAR and hyperspectral data detailed in Section 2.1 contain information about the structure and reflectance of the forest canopy. As noted in Section 1, this information has been shown to explain some of the variability in the forest variables of interest. Following post-processing, 153 and 224 values defined each LiDAR and hyperspectral signal, respectively. Obviously, we cannot include all of these highly collinear covariates in a regression model, and thus must consider some way to reduce the signals' dimension. Several variable extraction methods have been proposed for these types of data, see, e.g., Næsset and Gobakken (2008).

Our aim is to extract the maximum amount of information from the LiDAR and hyperspectral signals using the fewest number of non-collinear derived covariates. To be specific, we form the $u \times v$ matrix M with rows corresponding to signal measurement locations and columns holding signal values that have been centered (e.g., for the LiDAR data M is 12,414×153). Let $M = UDV'$ be the Singular Value Decomposition (SVD) of M, where U and V are orthogonal matrices of dimension $u \times u$ and $v \times v$ respectively, and $D = \begin{bmatrix} \Lambda \\ O \end{bmatrix}$, where Λ is a $v \times v$ diagonal matrix with diagonal entries $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_v \geq 0$ called the singular values of M and O is a $(u - v) \times v$ matrix of zeroes. In practice, one could look at $\mathrm{Tr}(M'M) = \sum_{i=1}^{v} \lambda_i^2$ as a measure of total variation and choose $r < v$ such that $\sum_{i=1}^{r} \lambda_i^2 / \sum_{i=1}^{v} \lambda_i^2 \approx 1$. In other words, $\lambda_i \approx 0$ for $i = r+1, r+2, \ldots, v$. We now form the matrix of "derived covariates" as $Z = MV_1$, where V_1 is the matrix formed from the first r columns of V. From the orthogonality of V, it is easy to see that $MV = UD$, which implies that $Z = MV_1 = U_1 \Lambda$, where U_1 is the matrix formed from the first r columns of U. This shows that $\boldsymbol{z}_i = \lambda_i \boldsymbol{u}_i = M\boldsymbol{v}_i$ for $i = 1, 2, \ldots, r$, where $\boldsymbol{u}_i$ is the ith column of U. These reveal that the derived covariates are orthogonal vectors (since the $\boldsymbol{u}_i$'s are) and they all belong to the column space of M – which is same as the original covariate space.

This method was applied to both the LiDAR and the hyperspectral data. Exploratory data analysis was used to determine the number of covariates to retain from each dataset. Specifically, we used non-spatial univariate regression models to explore a range of covariates for each outcome variable. The results showed that no covariates beyond the first four were significant in the models. The first four covariates explained 64% and 92% of the variability in the LiDAR and hyperspectral data, respectively, and were used for the subsequent analysis.

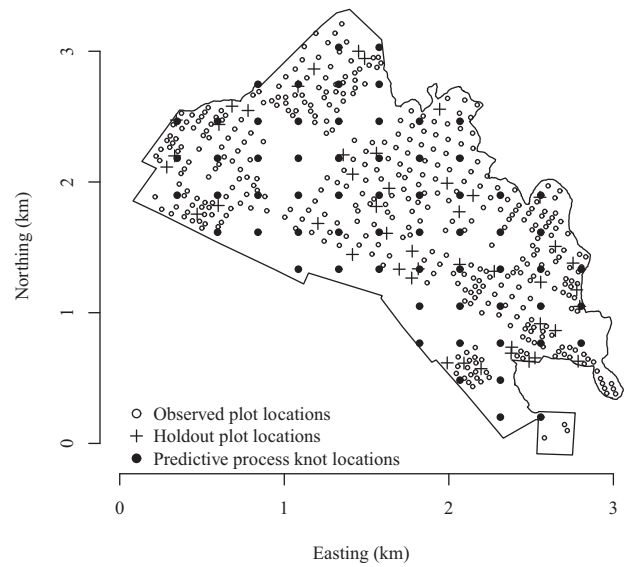

Fig. 2. Observed forest inventory plot locations along with the model validation holdout plot subset and the locations of the predictive process knots used in the 63 knot model.

3. Model implementation and analysis

Predictive process knots used for the multivariate and univariate bias-corrected predictive process versions of (1) were laid out on a grid over the PEF as illustrated in Fig. 2. For the subsequent analysis, we considered knot intensities of 63 and 124. In practice a range of knot intensities and configurations should be considered to explore the robustness of the model parameter estimates, see, e.g., Guhaniyogi et al. (2011) for discussion.

Prior to regression analysis, BIO and BA were square root transformed and DEN, DBH, and GRO were log transformed to better approximate a normal distribution and hence satisfy the models' assumed distribution of the outcome variable.

As detailed in Section 2.2.2, model fit was evaluated using *DIC*, *D*, and R^2 criteria. Predictive performance was assessed using RMSPE calculated using 46 randomly selected observed locations, i.e., a 10% holdout set, '+' symbol in Fig. 2. Candidate and baseline model parameters were estimated using the remaining 405 observations.

Parameter and predictive inference was based on three chains of 50,000 iterations (the first 10,000 iterations were discarded as burn-in). The marginalized sampler detailed in Section 2.2 was coded in C++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 Nahalem hyperthreaded quad-Xeon processors. For the spatial models, sampling the three chains of 50,000 iterations required approximately 10 and 16 h for the 63 and 124 knot models to complete.

To complete the proposed models' Bayesian specification, prior distributions were assigned to each parameter. As is customary, a flat prior was assigned to the regression intercept and slope parameters. The covariance matrices AA' and Ψ, were each assigned an $IW(a, b)$ with the degrees of freedom, a, set to $m + 1$. For AA' and Ψ, the IW scale matrix, b, was constructed with zeros on the

Table 1
Regression slope parameter credible intervals, 50 (2.5, 97.5) percentiles, for the multivariate candidate models.

Parameter	Non-spatial	Spatial	
		63 knot	124 knot
$\beta_{BIO,0}$	**10.03 (9.84, 10.22)**	**10.46 (9.25, 11.93)**	**10.22 (8.92, 11.41)**
$\beta_{BIO,1}$	**−158.48 (−181.48, −135.83)**	**−123.45 (−140.71, −105.29)**	**−115.82 (−132.96, −99.55)**
$\beta_{BIO,2}$	**−84.32 (−107.97, −59.82)**	**−79.28 (−99.24, −57.86)**	**−76.31 (−92.01, −58.90)**
$\beta_{BIO,3}$	−4.26 (−24.50, 20.54)	11.58 (−7.89, 32.42)	14.56 (−9.24, 34.81)
$\beta_{BIO,4}$	6.71 (−16.09, 26.76)	0.85 (−14.85, 16.73)	−2.02 (−15.61, 15.81)
$\beta_{BIO,5}$	**−29.13 (−50.43, −8.55)**	−14.41 (−35.65, 9.26)	−15.15 (−36.13, 4.14)
$\beta_{BIO,6}$	13.83 (−10.37, 36.17)	−13.66 (−32.86, 8.78)	−17.66 (−38.52, 5.49)
$\beta_{BIO,7}$	**−28.98 (−54.09, −3.81)**	−1.77 (−20.72, 17.54)	−5.19 (−24.59, 14.10)
$\beta_{BIO,8}$	**−27.40 (−49.02, −7.53)**	−9.91 (−31.08, 9.33)	−13.04 (−35.00, 5.25)
$\beta_{BA,0}$	**4.85 (4.75, 4.95)**	**5.05 (4.48, 5.80)**	**4.95 (4.30, 5.52)**
$\beta_{BA,1}$	**−63.17 (−74.64, −51.55)**	**−47.17 (−55.62, −38.94)**	**−43.59 (−52.18, −36.01)**
$\beta_{BA,2}$	**−9.74 (−20.75, 2.96)**	**−14.31 (−24.30, −4.82)**	**−13.06 (−20.44, −4.20)**
$\beta_{BA,3}$	−4.06 (−15.44, 7.84)	3.30 (−5.84, 13.77)	5.71 (−5.58, 15.96)
$\beta_{BA,4}$	**11.46 (1.27, 21.60)**	6.51 (−0.73, 13.41)	5.55 (−2.86, 13.91)
$\beta_{BA,5}$	−5.09 (−16.76, 5.38)	−2.12 (−12.70, 9.51)	−2.92 (−13.70, 6.86)
$\beta_{BA,6}$	4.90 (−8.02, 17.11)	−7.37 (−17.47, 2.80)	−8.26 (−19.17, 2.89)
$\beta_{BA,7}$	**−12.55 (−25.73, −0.21)**	0.71 (−7.20, 10.67)	−0.22 (−9.53, 7.91)
$\beta_{BA,8}$	−10.55 (−21.41, 0.20)	−4.33 (−13.47, 4.67)	−6.45 (−15.59, 2.05)
$\beta_{DEN,0}$	**8.14 (8.02, 8.25)**	**8.10 (7.29, 8.95)**	**8.01 (7.07, 8.66)**
$\beta_{DEN,1}$	−5.08 (−17.64, 8.02)	2.52 (−6.25, 12.19)	4.20 (−4.81, 11.43)
$\beta_{DEN,2}$	**41.02 (27.55, 55.73)**	**31.40 (21.16, 39.98)**	**32.00 (23.78, 41.38)**
$\beta_{DEN,3}$	−4.58 (−17.90, 9.99)	**10.49 (1.20, 18.19)**	**10.76 (2.64, 20.36)**
$\beta_{DEN,4}$	1.49 (−10.70, 14.77)	4.46 (−3.31, 13.79)	3.74 (−3.74, 12.67)
$\beta_{DEN,5}$	**12.21 (0.03, 27.40)**	2.31 (−8.20, 12.50)	0.35 (−10.34, 10.92)
$\beta_{DEN,6}$	**23.08 (8.06, 36.25)**	4.52 (−6.57, 14.12)	3.77 (−8.70, 14.16)
$\beta_{DEN,7}$	**−17.23 (−32.36, −3.75)**	0.76 (−8.59, 8.59)	−1.17 (−10.58, 8.18)
$\beta_{DEN,8}$	−5.65 (−18.15, 8.41)	0.05 (−8.67, 10.06)	−1.59 (−10.64, 8.54)
$\beta_{DBH,0}$	**1.55 (1.52, 1.59)**	**1.61 (1.35, 1.87)**	**1.60 (1.36, 1.89)**
$\beta_{DBH,1}$	**−14.09 (−18.23, −10.16)**	**−12.83 (−16.20, −8.57)**	**−12.32 (−15.67, −9.35)**
$\beta_{DBH,2}$	**−22.60 (−27.16, −18.29)**	**−18.40 (−22.09, −13.93)**	**−18.13 (−22.71, −14.64)**
$\beta_{DBH,3}$	−2.17 (−6.66, 1.96)	**−5.09 (−9.00, −1.10)**	**−5.22 (−8.50, −1.86)**
$\beta_{DBH,4}$	0.35 (−3.90, 4.24)	−1.30 (−4.56, 1.58)	−1.39 (−4.37, 2.00)
$\beta_{DBH,5}$	**−9.47 (−13.72, −5.40)**	−3.22 (−7.34, 0.89)	−2.34 (−7.35, 2.11)
$\beta_{DBH,6}$	−4.50 (−8.66, 0.05)	−1.89 (−6.05, 1.99)	−2.07 (−5.79, 2.68)
$\beta_{DBH,7}$	3.62 (−0.64, 7.85)	−0.09 (−4.13, 3.98)	−0.07 (−3.69, 3.75)
$\beta_{DBH,8}$	−1.03 (−5.41, 3.15)	−0.15 (−4.04, 3.56)	0.43 (−3.67, 4.28)
$\beta_{GRO,0}$	**1.26 (1.20, 1.31)**	**1.23 (0.99, 1.40)**	**1.21 (0.99, 1.45)**
$\beta_{GRO,1}$	**−10.30 (−16.17, −4.77)**	**−11.37 (−16.22, −5.02)**	**−11.04 (−17.06, −5.86)**
$\beta_{GRO,2}$	3.83 (−2.24, 9.53)	4.89 (−2.43, 10.77)	4.97 (−1.16, 11.60)
$\beta_{GRO,3}$	**8.95 (3.02, 15.81)**	**11.27 (5.34, 17.05)**	**10.55 (4.13, 16.18)**
$\beta_{GRO,4}$	−3.61 (−8.54, 1.57)	−1.56 (−6.50, 4.07)	−0.99 (−6.36, 3.63)
$\beta_{GRO,5}$	**12.64 (7.42, 18.34)**	4.58 (−1.15, 10.53)	4.27 (−2.15, 11.37)
$\beta_{GRO,6}$	**6.75 (0.03, 12.55)**	4.58 (−2.98, 11.54)	5.47 (−1.68, 12.43)
$\beta_{GRO,7}$	2.75 (−3.09, 8.42)	3.97 (−1.45, 9.40)	4.53 (−1.62, 9.71)
$\beta_{GRO,8}$	4.35 (−1.02, 9.80)	−0.62 (−6.37, 4.91)	−1.12 (−7.75, 4.03)

off-diagonal elements and diagonal elements taken as the partial-sill and nugget values, respectively, from univariate semivariograms fit to the residuals of the non-spatial multivariate model. These semivariograms are illustrated in Fig. 3. To reduce the computational complexity, we fixed $\nu = 0.5$ which reduces (2) to an exponential correlation function. Fixing the Matérn smoothness parameter does reduce flexibility; however, this concession results in substantial reduction in computing time. The spatial decay parameter ϕ, for each of the m correlation functions, were assigned a uniform $U(0.001, 0.3)$, which, given $\nu = 0.5$, corresponds to support for an *effective spatial range* between approximately 10 and 3000 m (i.e., where effective spatial range is the distance at which the correlation drops to 0.05, $-\log(0.05)/\phi$).

4. Results and discussion

Parameter estimates for the multivariate spatial and non-spatial regression models were obtained following the methods detailed in Section 3. Results suggested that none of the covariates derived from the hyperspectral data and few from the LiDAR data explained a statistically significant amount of variability in the outcomes. The covariates' regression slope parameter estimates are provided in Table 1. Here, subscripts on the slope parameters (β's) indicate the outcome variable and associated covariate. The $\beta_{\cdot,0}$ correspond to the intercept whereas $\beta_{\cdot,1} \cdots \beta_{\cdot,4}$ and $\beta_{\cdot,5} \cdots \beta_{\cdot,8}$ correspond to the first four LiDAR and hyperspectral derived covariates, respectively. Values in bold indicate those slope parameters with 95% credible intervals that do not include zero. There were several statistically significant slope parameters in the non-spatial model that were not significant in the spatial models – particularly those parameters associated with the hyperspectral covariates. This is a result of violating the key model assumption of independent and identically distributed residuals. The violation of this assumption is clearly seen in Fig. 3 where outcome specific semivariograms are fit to the non-spatial model's residuals.

Given these initial results, we refit the regression models using only those covariates that had statistically significant slope parameters in the multivariate spatial regression models. The resulting parameter estimates are provided in Table 2. Although not shown, the univariate spatial regression models slope parameter estimates did not differ substantially from those of the multivariate spatial regression models.

The model fit and predictive performance of the benchmark and candidate models are summarized in Table 3. Considering first the

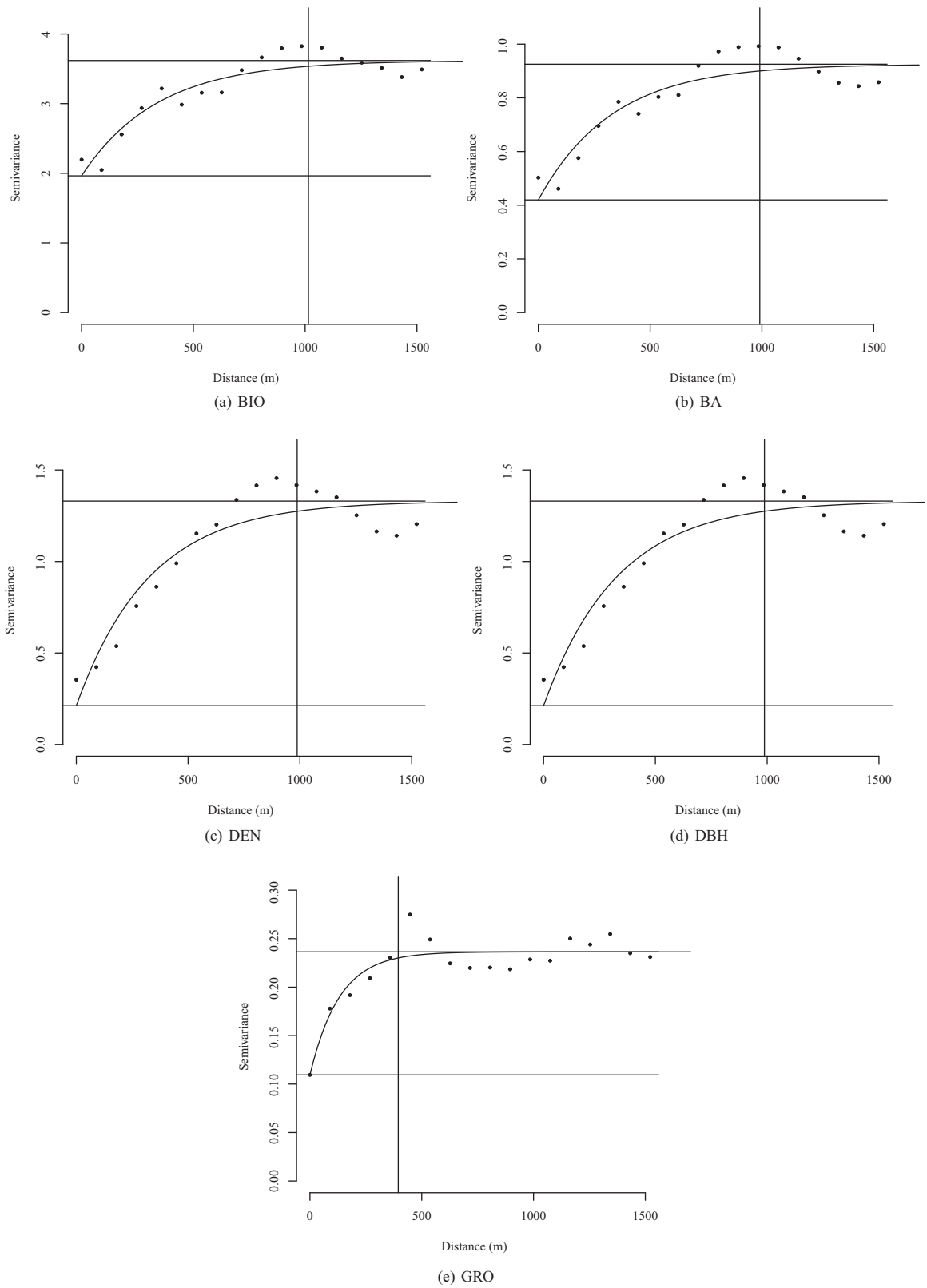

Fig. 3. Empirical semivariograms of residuals from non-spatial version of model (1) along with the maximum likelihood estimates of the exponential semivariogram model function's nugget (lower horizontal line), sill (upper horizontal line), and effective spatial range (vertical line).

results for the multivariate models (columns 4–6 in Table 3); compared to the non-spatial specification, the spatial models produce considerably lower *DIC* and *D*, which suggests the addition of the multivariate spatial random effects improve model fit. Similarly, the outcome variable specific R^2 are substantially higher among the spatial models compared to the non-spatial model. The multivariate spatial models also produce consistently lower RMSPE compared to their non-spatial counterpart. Among the spatial

Table 2
Regression slope parameter credible intervals, 50 (2.5, 97.5) percentiles, for the multivariate candidate models.

Parameter	Non-spatial	Spatial	
		63 knot	124 knot
$\beta_{BIO,0}$	10.05 (9.87, 10.24)	10.45 (9.39, 11.56)	10.12 (8.49, 12.18)
$\beta_{BIO,1}$	−168.15 (−185.87, −153.66)	−123.88 (−136.96, −108.62)	−117.32 (−131.53, −98.55)
$\beta_{BIO,2}$	−100.68 (−121.65, −76.94)	−82.72 (−102.26, −69.25)	−79.63 (−94.03, −61.98)
$\beta_{BA,0}$	4.86 (4.77, 4.96)	5.07 (4.48, 5.55)	4.88 (3.94, 5.86)
$\beta_{BA,1}$	−65.44 (−72.61, −59.06)	−47.68 (−53.77, −40.66)	−44.59 (−51.84, −37.68)
$\beta_{BA,2}$	−18.00 (−29.20, −6.31)	−17.80 (−25.69, −10.34)	−16.81 (−24.52, −9.17)
$\beta_{DEN,0}$	8.10 (7.99, 8.20)	8.09 (7.46, 8.76)	8.04 (6.97, 8.98)
$\beta_{DEN,2}$	34.29 (23.86, 46.44)	26.92 (18.30, 34.87)	25.96 (18.21, 34.70)
$\beta_{DEN,3}$	−4.76 (−13.34, 4.06)	11.37 (2.06, 20.38)	9.30 (3.01, 15.01)
$\beta_{DBH,0}$	1.57 (1.54, 1.60)	1.64 (1.39, 1.83)	1.61 (1.32, 1.94)
$\beta_{DBH,1}$	−17.19 (−19.90, −14.97)	−11.70 (−13.19, −9.66)	−10.94 (−13.27, −8.92)
$\beta_{DBH,2}$	−22.18 (−26.06, −18.52)	−16.83 (−20.66, −13.65)	−16.06 (−19.89, −13.23)
$\beta_{DBH,3}$	−2.26 (−5.68, 0.99)	−6.68 (−10.94, −2.63)	−6.16 (−9.40, −3.54)
$\beta_{GRO,0}$	1.24 (1.20, 1.29)	1.22 (0.95, 1.49)	1.21 (0.96, 1.42)
$\beta_{GRO,1}$	−6.20 (−11.12, −1.53)	−12.45 (−17.10, −6.06)	−12.73 (−17.37, −8.23)
$\beta_{GRO,3}$	6.86 (1.03, 12.56)	9.90 (4.07, 14.37)	9.29 (4.35, 13.67)

models, increasing n^* should result in improved fit and predictive performance. This result is expected because a larger number of knots provide a better approximation of the parent process. This improved fit to the data is seen between the $n^* = 63$ and 124 knot models (i.e., the 124 knot model has lower *DIC* and *D*, and larger R^2 values). However, the 124 knot model does not consistently outperform the 63 knot model when considering the holdout set RMSPE (e.g., RMSPE_{BIO} is 27.57 and 28.40 for $n^* = 63$ and 124, respectively). Such variation is expected because the parent process is being estimated with a different, albeit marginally lower intensity, knot configuration.

Comparing between fit criteria and RMSPE for multivariate (or univariate) non-spatial and spatial models in Table 3 suggest the LiDAR covariates were not very useful for explaining variability in the outcomes. Although LiDAR data are able to quantify some vertical and horizontal forest structure characteristics, the substantial gains in fit and predictive performance are realized through the addition of the spatial random effects.

The first column in Table 3 provides the fit and holdout set RMSPE for the *k*-NN benchmark. These results are based on a $k = 15$ and the set of covariates considered in Table 2. The $k = 15$ was chosen because it resulted in the lowest RMSPE for the majority of the outcomes (specifically for BIO, BA, and DBH). Compared to the multivariate spatial models, *k*-NN provided lower R^2 but similar values of RMSPE. We note, the *k*-NN algorithm could produce different, and perhaps improved results, if one pursued an exhaustive cross-validation approach for choosing the covariates and associated weighting for defining an *optimal* NN search space. This however is beyond the scope of the current study.

We are also interested in assessing if the complexity of the multivariate spatial model improves fit and predictive performance, compared to results obtained using simpler regression models. Comparison between the univariate and multivariate 124 knot model R^2 and RMSPE values, columns 3 and 6 respectively, shows that with the exception of tree density, DEN, results are comparable and in some instances the univariate model provides marginally better fit and prediction. For DEN, however, the multivariate model provides substantially improved model R^2 and RMSPE. These improvements are likely due to explicit modeling of the covariance among the outcome variables' spatial random effects through the cross-covariance matrix AA'. Estimates of these and the other parameters associated with the models' residual components are summarized in Table 4.

To aid in interpretation, we converted off-diagonal elements of the AA' covariance matrix to their corresponding correlations, e.g., the median of $AA'_{BIO,BA}$ correlation posterior distribution is $\rho_{BIO,BA} = 0.94$ in Table 4. Here, the cross-correlation

estimates in bold have 95% credible intervals that do not include zero. By considering the sign and strength of these correlation estimates, along with the scatter plot of the

Table 3
Benchmark and candidate model fit and predictive performance.

	k-NN ($k = 15$)	Univariate		Multivariate		
		Non-spatial	Spatial	Non-spatial	Spatial	
			$n^* = 124$		$n^* = 63$	$n^* = 124$
pD	–	–	–	28	176	258
DIC	–	–	–	−895	−1888	−1941
G	–	–	–	2651	1095	957
P	–	–	–	2555	104	139
D	–	–	–	5206	1199	1095
R^2_{BIO}	0.55	0.42	0.73	0.40	0.71	0.73
R^2_{BA}	0.46	0.28	0.70	0.28	0.67	0.70
R^2_{DEN}	0.32	0.08	0.52	0.08	0.72	0.75
R^2_{DBH}	0.37	0.30	0.69	0.26	0.61	0.63
R^2_{GRO}	0.17	0.01	0.38	0.02	0.37	0.38
RMSPE_{BIO}	27.91	32.28	26.58	31.82	27.57	28.40
RMSPE_{BA}	6.58	7.12	6.11	7.06	6.29	6.55
RMSPE_{DEN}	4288.84	4532.38	4153.61	4465.55	4024.50	3874.31
RMSPE_{DBH}	1.67	1.72	1.23	1.72	1.20	1.24
RMSPE_{GRO}	1.67	1.64	1.11	1.66	1.14	1.11

Table 4
Regression spatial and residual parameter credible intervals, 50 (2.5, 97.5) percentiles, for the multivariate candidate models.

Parameter	Non-spatial	Spatial	
		63 knot	124 knot
$AA'_{BIO,BIO}$	–	3.33 (2.31, 4.02)	4.77 (2.88, 6.00)
$AA'_{BA,BA}$	–	0.78 (0.61, 0.92)	1.22 (0.73, 1.48)
$AA'_{DEN,DEN}$	–	0.90 (0.77, 1.13)	1.53 (1.09, 1.96)
$AA'_{DBH,DBH}$	–	0.12 (0.10, 0.17)	0.18 (0.12, 0.23)
$AA'_{GRO,GRO}$	–	0.19 (0.12, 0.27)	0.17 (0.11, 0.26)
$\rho_{BIO,BA}$	–	**0.94 (0.90, 0.95)**	**0.94 (0.88, 0.96)**
$\rho_{BIO,DEN}$	–	**0.46 (0.38, 0.58)**	**0.59 (0.42, 0.67)**
$\rho_{BIO,DBH}$	–	0.09 (−0.07, 0.21)	−0.08 (−0.22, 0.09)
$\rho_{BIO,GRO}$	–	−0.11 (−0.43, 0.21)	−0.08 (−0.45, 0.19)
$\rho_{BA,DEN}$	–	**0.64 (0.54, 0.70)**	**0.75 (0.67, 0.81)**
$\rho_{BA,DBH}$	–	−0.13 (−0.28, 0.02)	**−0.32 (−0.47, −0.08)**
$\rho_{BA,GRO}$	–	0.01 (−0.28, 0.26)	0.07 (−0.37, 0.34)
$\rho_{DEN,DBH}$	–	**−0.74 (−0.80, −0.67)**	**−0.81 (−0.86, −0.65)**
$\rho_{DEN,GRO}$	–	**0.16 (0.01, 0.35)**	**0.36 (0.07, 0.55)**
$\rho_{DBH,GRO}$	–	**−0.26 (−0.62, −0.01)**	**−0.52 (−0.65, −0.29)**
$\Psi_{BIO,BIO}$	3.63 (3.29, 4.20)	1.00 (0.86, 1.35)	1.02 (0.78, 1.19)
$\Psi_{BIO,BA}$	1.72 (1.55, 1.98)	0.45 (0.38, 0.61)	0.45 (0.33, 0.52)
$\Psi_{BIO,DEN}$	1.25 (1.15, 1.47)	0.11 (−0.00, 0.20)	0.02 (−0.06, 0.11)
$\Psi_{BIO,DBH}$	−0.05 (−0.09, 0.00)	0.06 (0.01, 0.10)	0.08 (0.05, 0.13)
$\Psi_{BIO,GRO}$	0.22 (0.12, 0.30)	0.22 (0.10, 0.30)	0.20 (0.12, 0.26)
$\Psi_{BA,BA}$	0.91 (0.83, 1.05)	0.25 (0.20, 0.31)	0.23 (0.17, 0.27)
$\Psi_{BA,DEN}$	0.78 (0.72, 0.88)	0.12 (0.06, 0.16)	0.05 (0.00, 0.09)
$\Psi_{BA,DBH}$	−0.10 (−0.12, −0.07)	0.00 (−0.02, 0.02)	0.02 (0.00, 0.04)
$\Psi_{BA,GRO}$	0.15 (0.09, 0.19)	0.12 (0.07, 0.15)	0.11 (0.07, 0.14)
$\Psi_{DEN,DEN}$	1.30 (1.18, 1.42)	0.23 (0.19, 0.27)	0.14 (0.11, 0.18)
$\Psi_{DEN,DBH}$	−0.33 (−0.37, −0.29)	−0.08 (−0.10, −0.06)	−0.05 (−0.07, −0.04)
$\Psi_{DEN,GRO}$	0.21 (0.16, 0.26)	0.08 (0.05, 0.11)	0.06 (0.03, 0.09)
$\Psi_{DBH,DBH}$	0.13 (0.12, 0.15)	0.05 (0.04, 0.06)	0.04 (0.03, 0.05)
$\Psi_{DBH,GRO}$	−0.07 (−0.08, −0.05)	−0.01 (−0.03, −0.00)	−0.01 (−0.02, 0.00)
$\Psi_{GRO,GRO}$	0.26 (0.23, 0.30)	0.12 (0.10, 0.15)	0.12 (0.08, 0.16)
Eff. range_1 (m)	–	1360.97 (804.22, 1709.15)	1727.45 (1012.69, 1927.03)
Eff. range_2 (m)	–	1931.54 (1175.13, 1987.30)	1607.31 (1259.96, 1903.29)
Eff. range_3 (m)	–	1808.65 (1329.55, 1969.88)	1607.95 (1136.12, 1915.34)
Eff. range_4 (m)	–	1920.95 (1497.45, 1995.78)	1953.56 (1864.27, 1991.62)
Eff. range_5 (m)	–	1071.85 (695.11, 1547.14)	737.09 (424.88, 1280.27)

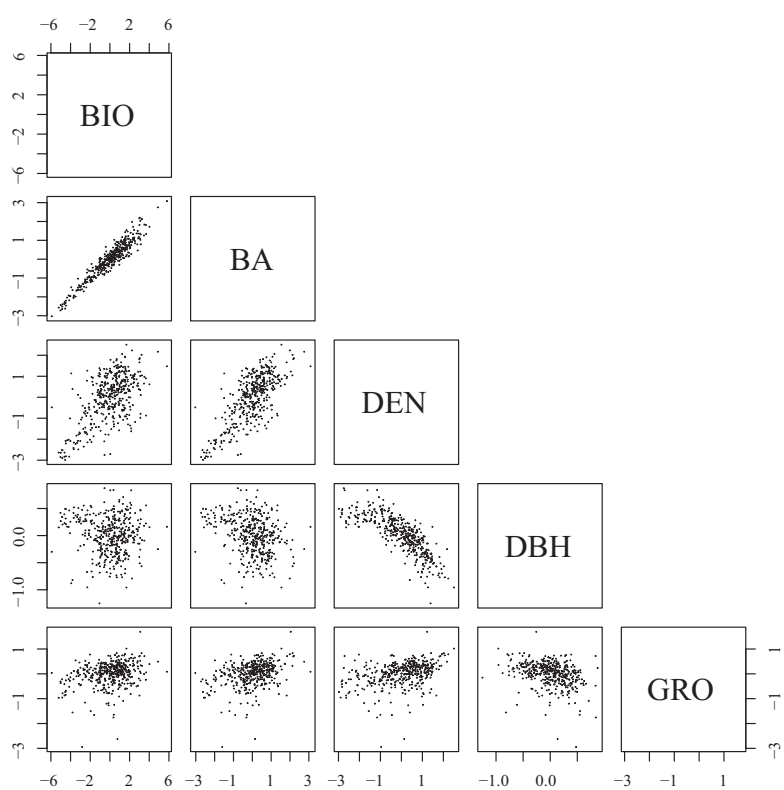

Fig. 4. Scatter plots of residuals from non-spatial version of the multivariate candidate model.

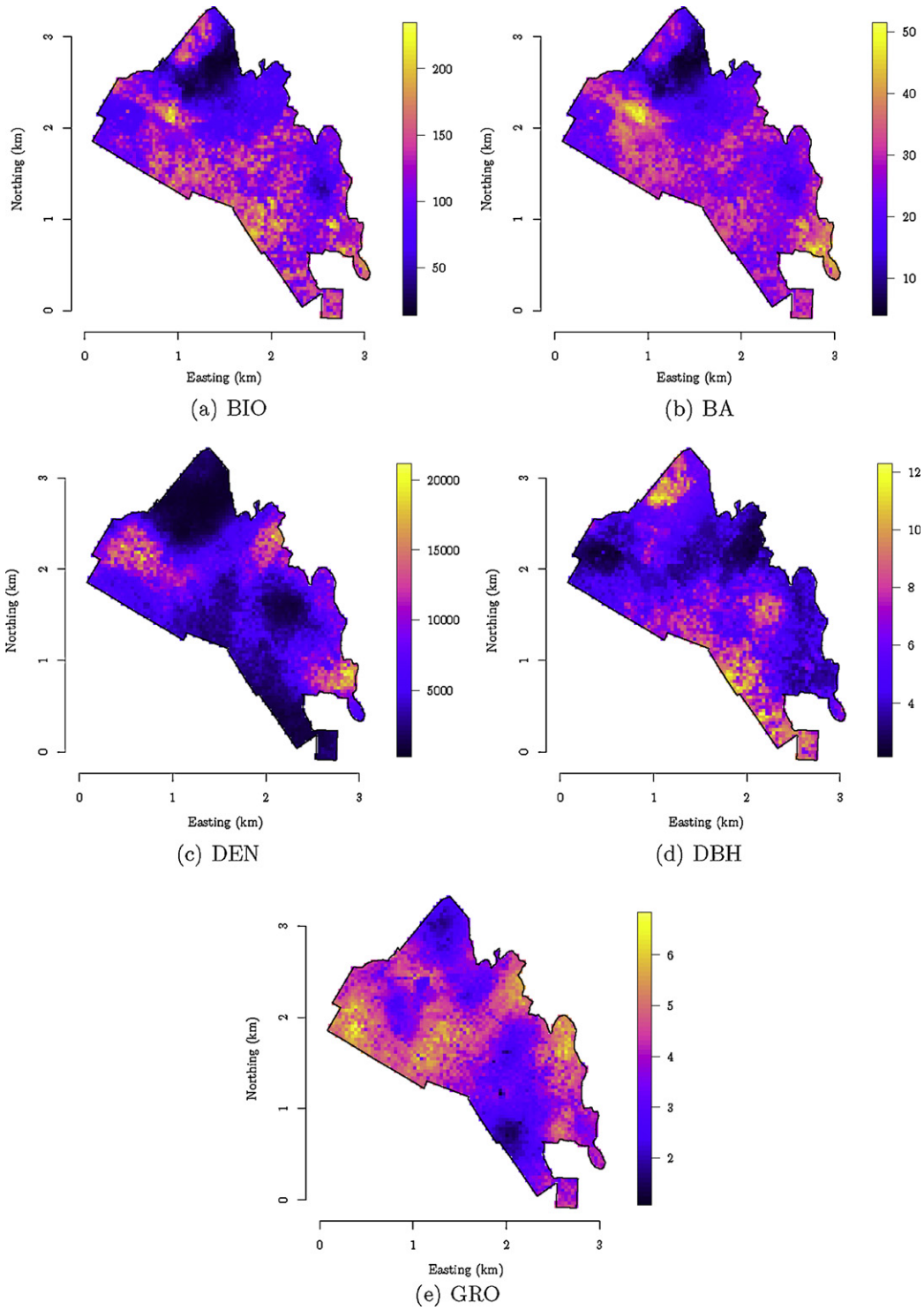

Fig. 5. Mean of the posterior predicted distribution at the pixel-level for each outcome variable.

Table 5
Predicted total and standard deviation in parentheses of the outcome variables over the 434.1 ha PEF study area.

	Design-based	k-NN (k = 15)	Multivariate		
			Non-spatial	Spatial	
				68 knot	124 knot
BIO	43,801.84 (903.06)	45,899.08	46,875.98 (884.43)	46,512.13 (968.07)	46,279.79 (954.89)
BA	10,427.92 (206.38)	10,603.97	10,853.98 (216.97)	10,695.81 (213.96)	10,661.29 (219.83)
DEN	2,541,714 (103,402.4)	1,710,907	2,936,691 (196,664.5)	2,309,598 (117,339.5)	2,327,825 (112,891.5)
DBH	4.94 (2.01)	4.88	5.28 (0.10)	5.54 (0.14)	5.51 (0.13)
GRO	1662.72 (35.49)	1560.728	1723.47 (46.29)	1691.72 (49.84)	1684.54 (47.20)

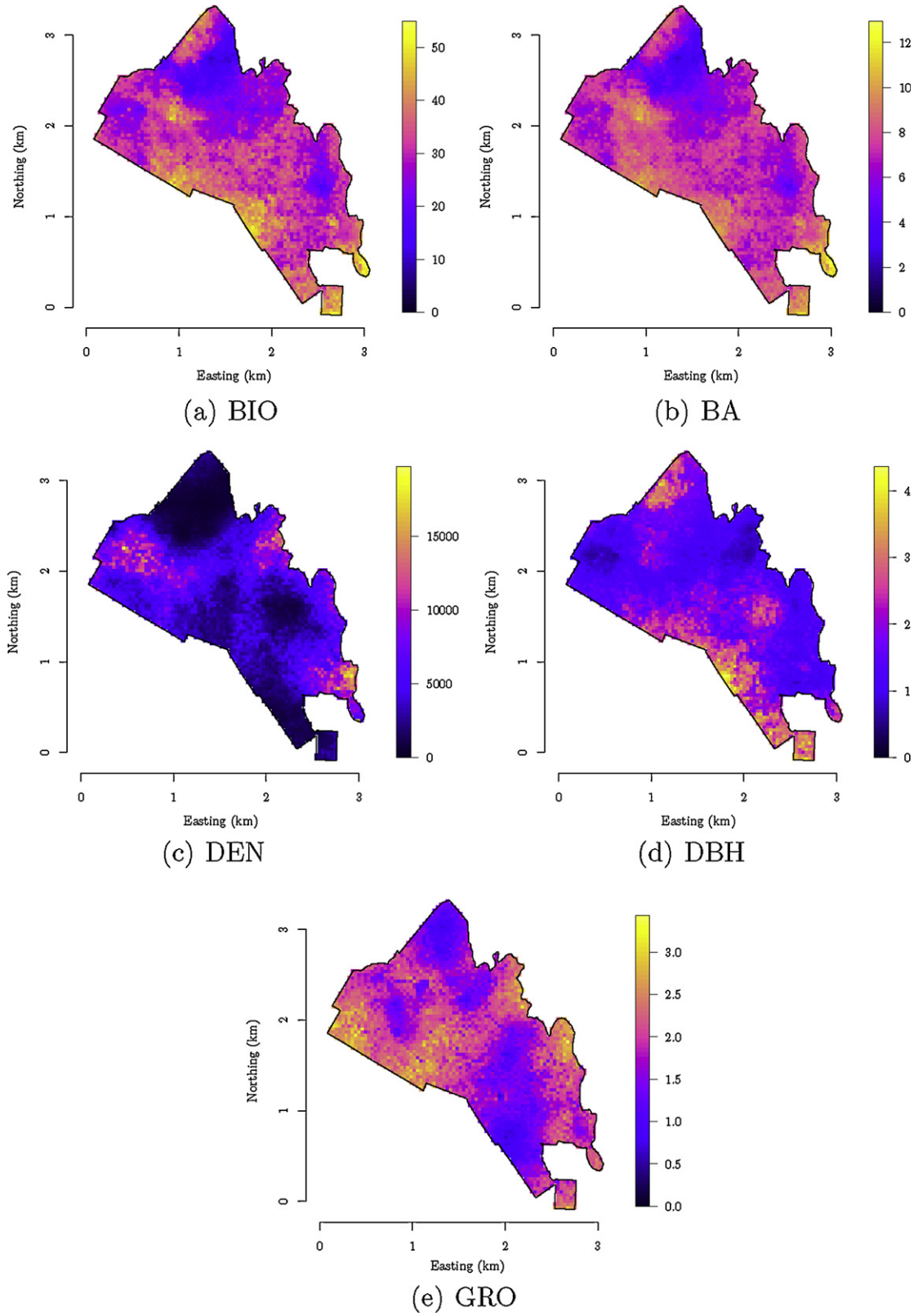

Fig. 6. Standard deviation of the posterior predicted distribution at the pixel-level for each outcome variable.

non-spatial model's residuals, Fig. 4, it becomes clear that the co-varying spatial random effects improve model fit and subsequent prediction for outcomes, such as DEN, that are poorly quantified by the LiDAR covariates. An additional advantage to using the multivariate spatial random effects is that the outcome variables predicted at a new location maintain a realistic covariance that might otherwise be lost when outcomes are modeled independent of each other.

As noted above, the multivariate and univariate spatial regression models effectively leverage the dependence among

observed outcomes to improve fit and prediction. Residual spatial dependence extends well beyond 1 km for all outcome variables, as seen in the exploratory data analysis semivariograms, Fig. 3, and in the estimated effective spatial ranges of the spatial random effects in Table 4. Here, the disparity among the effective ranges supports our use of a non-separable specification of the spatial cross-covariance function.

Fig. 5 shows maps comprising 12,414 pixels' posterior predictive distribution mean generated using the 63 knot multivariate spatial model. As suggested by the holdout set RMSPE results, there was negligible difference between maps based on the 63 and 124 knot models. These maps are comparable to the patterns seen in the observed data, Fig. 1. Importantly, because we used a Bayesian inferential framework, we have access to the entire posterior predictive distribution at all new locations (or pixels). This flexibility allows us to create maps of any function of the predictive distribution. For example, creating maps of the untransformed predicted outcomes in Fig. 5 simply requires applying the inverse transformation to each posterior predictive sample then taking the mean over all resulting sample values. To assess prediction uncertainty we might look at the range between each pixel's upper and lower 95% percentile or perhaps standard deviation of the posterior predictive distribution samples as illustrated in Fig. 6. Such maps are useful for designing future forest inventory protocol.

In addition to small area prediction, e.g., at the pixel-level, there is often interest at summarizing the outcome variable at some broader management unit. Table 5 provides totals and associated standard deviation for each outcome variable over the PEF. The first column labeled "Design-based" are totals and standard errors estimated using a finite population approach that assumes the observations were drawn using simple random sampling (see, e.g., Gregoire and Valentine, 2008). The second column provides results obtained using the *k*-NN algorithm. Although not pursued here, McRoberts et al. (2007) detail approaches for deriving standard errors for the *k*-NN based total. The results in columns 3–5 are calculated using the posterior predictive distributions from the multivariate non-spatial and spatial models. When considering the estimated uncertainty, there are negligible differences among the totals produced using these various approaches.

5. Conclusions

In this paper we compared multivariate regression models with and without spatially structured random effects to several alternative, less computationally, complex models. In addition to helping meet the regression model's assumption of independent and identically distributed residuals, the results of the PEF data analysis suggest that the addition of univariate or multivariate spatial random effects greatly improved model fit and prediction. In some cases, explicitly modeling the residual spatial dependence among the outcomes via multivariate random effects resulted in improved inference, e.g., as in the case of the tree density outcome.

For the spatial regression models, we achieve computational efficiency by replacing the parent spatial process by its induced predictive process. The appeal of this solution is that one need not digress from the modeling objectives to think about choices of basis functions, or kernels or alignment algorithms for the locations. The analysis demonstrated that a low-dimensional representation of the parent process is highly effective at borrowing information over space to deliver accurate and precise prediction at new locations. Further, by working in a Bayesian paradigm we have access to the full posterior predictive distribution at each new location, which facilitates uncertainty assessment. Access to posterior predictive samples could also be useful in settings where the investigator wishes to propagate uncertainty in outcome variables through subsequent economic or ecological numerical models. For example, many forest yield models require spatially and temporally explicit inputs that must be somehow imputed to the region of interest.

Variable extraction from the high-dimensional LiDAR and hyperspectral data was achieved using a straightforward approach that maximizes signal information using the fewest number of non-collinear derived covariates. Although several studies have used AVIRIS data to partially explain indices of forest health or productivity, such as foliar nitrogen concentration Martin et al. (2008), our derived covariates did not explain much variability in annual growth or other outcome variables. Although several LiDAR covariates were found to be useful, they explained a relatively small amount of variability in the outcomes as reflected in the low R^2 and holdout set RMSPE values.

There are obviously many alternative approaches to deriving covariates from these data, some of which might result in more useful covariates, i.e., those that explain more variability in the outcome variables. For instance, Lefsky et al. (2002) demonstrate that LiDAR derived covariates related to maximum canopy height are strongly correlated with several important forest variables. Such covariates could certainly be included in the regression model considered in our analysis. However, in practice the addition of spatial random effects almost always results in a gain in model fit and predictive ability. Further, as demonstrated by the comparison between the non-spatial and spatial random effects models, the presence of correlated residuals can result in misleading inference regarding the importance of the covariates.

Our future research will focus on developing modeling frameworks that automate the selection of covariates from high-dimensional remotely sensed data. Also, we are considering application of non-stationary cross-covariance functions that allow covariance among outcomes to vary by location.

Acknowledgments

The work of the first author was supported by the USDA Forest Service, Forest Inventory and Analysis National Program, Forest Health Technology Enterprise Team, and National Science Foundation (NSF) grant EF-1137309. The first and second authors were supported by the grant NSF-DMS-1106609. The first and fourth authors were supported by USDA/NASA grant 10-JV-11242307-037.

Data for this study were provided by a unit of the Northern Research Station, U.S. Forest Service, located at the Penobscot Experimental Forest in Maine. Significant funding for collection of these data was provided by the U.S. Forest Service. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

We greatly appreciate the useful comments and suggestions made by the Guest Editor and two anonymous reviewers.

References

Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., Smith, M.-L., Dubayah, R.O., Hofton, M.A., Blair, J.B., 2008. Integrating waveform LiDAR with hyperspectral imagery for inventory of a northern temperate forest. Remote Sensing of Environment 112, 1856–1870.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC Press, Boca Raton, FL.

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial datasets. Journal of the Royal Statistical Society Series B 70, 825–848.

Banerjee, S., Johnson, G.A., 2006. Coregionalized single- and multi-resolution spatially-varying growth curve modelling with application to weed growth. Biometrics 61, 617–625.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall/CRC Press, New York.

Cressie, N.A.C., 1993. Statistics for Spatial Data, 2nd edition. New York, Wiley.

Cressie, N., Johannesson, G., 2007. Spatial prediction for massive datasets. Journal of the Royal Statistical Society Series B 70, 209–226.

Crainiceaniu, C.M., Diggle, P.J., Rowlingson, B., 2008. Bivariate binomial spatial modeling of Loa Loa prevalence in Tropical Africa. Journal of the American Statistical Association 103, 21–37.
Daniels, M.J., Kass, R.E., 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. Journal of the American Statistical Association 94, 1254–1263.
Diggle, P., Lophaven, S., 2006. Bayesian geostatistical design. Scandinavian Journal of Statistics 33, 53–64.
Du, J., Zhang, H., Mandrekarm, V.S., 2009. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. The Annals of Statistics 37, 3330–3361.
Finley, A.O., Sang, H., Banerjee, S., Gelfand, A.E., 2009. Improving the performance of predictive process modeling for large datasets. Computational Statistics and Data Analysis 53, 2873–2884.
Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association 102, 321–331.
Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. Journal of Computational and Graphical Statistics 15, 502–523.
Gelfand, A.E., Ghosh, S.K., 1998. Model choice: a minimum posterior predictive loss approach. Biometrika 85, 1–11.
Gelfand, A.E., Schmidt, A., Banerjee, S., Sirmans, C.F., 2004. Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). Test 13, 1–50.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd edition. Chapman and Hall/CRC Press, Boca Raton, FL.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC Press, Boca Raton, FL.
Gregoire, T.G., Valentine, H.T., 2008. Sampling Strategies for Natural Resources and Environment. Chapman-Hall/CRC Press.
Grzebyk, M., Wackernagel, H., 1994. Multivariate analysis and spatial/temporal scales: real and complex models. In: Proceedings of the XVIIth International Biometrics Conference 19-33, Hamilton, Ontario.
Guhaniyogi, R., Finley, A.O., Banerjee, S., Gelfand, A.E., 2011. Adaptive Gaussian predictive process models for large spatial datasets. Environmetrics 22, 997–1007.
Higdon, D., 2002. Space and space-time modeling using process convolutions. In: Anderson, C., Barnett, V., Chatwin, P.C., El-Shaarawi, A.H. (Eds.), Quantitative Methods for Current Environmental Issues. Springer-Verlag.
Hoeting, J.A., 2009. The importance of accounting for spatial and temporal correlation in analyses of ecological data. Ecological Applications 19, 584–588.
Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2004. Comprehensive database of diameter-based biomass regressions for North American tree species. USDA Forest Service GTR319.
Kamman, E.E., Wand, M.P., 2003. Geoadditive models. Applied Statistics 52, 1–18.
Kaufman, C.G., Schervish, M.J., Nychka, D.W., 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. Journal of the American Statistical Association 103, 1545–1555.
Kamper, T.U., Johnson, B.R., Kuester, M., Keller, M., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. Journal of Applied Remote Sensing, http://dx.doi.org/10.1117/1.3361375.
Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. LiDAR remote sensing for ecosystem studies. Bioscience 52, 19–30.
Martin, M.E., Plourde, L.C., Ollinger, S.V., Smith, M.-L., McNeil, B.E., 2008. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. Remote Sensing of Environment 112, 3511–3519.
McRoberts, R.E., Tomppo, E., Finley, A.O., Heikkinen, J., 2007. Model-based estimators for the k-nearest neighbors technique. Remote Sensing of Environment 111, 466–480.
McRoberts, R.E., 2011. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management, http://dx.doi.org/10.1016/j.foreco.2011.06.039.
Næsset, E., Gobakken, T., 2008. Estimation of above and belowground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment 112, 3079–3090.
Nychka, D., Saltzman, N., 1998. Design of air-quality monitoring networks. In: Nychka, D., Cox, L., Piegorsch, W. (Eds.), Case studies in Environmental Statistics, Lecture Notes in Statistics 132. Springer Verlag, New York, pp. 51–76.
Ollinger, S.V., Richardson, A.D., Martin, M.E., Hollinger, D.Y., Frolking, S.E., Reich, P.B., Plourde, L.C., Katul, G.G., Munger, J.W., Oren, R., Smith, M.-L., Paw, U., Bolstad, K.T., Cook, P.V., Day, B.D., Martin, M.C., Monson, T.A., Schmid, R.K., 2008. Canopy nitrogen, carbon assimilation and albedo in temperate and boreal forests: functional relations and potential climate feedbacks. Proceedings of National Academy of Sciences of the United States of America 105, 19335–19340.
Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA.
Reich, B.J., Fuentes, M., 2007. A multivariate nonparametric Bayesian spatial framework for hurricane surface wind fields. The Annals of Applied Statistics 1, 249–264.
Robert, C.P., Casella, G., 2005. Monte Carlo Statistical Methods, 2nd edition. Springer, New York.
Schmidt, A., Gelfand, A.E., 2003. A Bayesian coregionalization model for multivariate pollutant data. Journal of Geophysics Research: Atmospheres 108, 8783.
Sendak, P.E., Brissette, J.C., Frank, R.M., 2003. Silviculture affects composition, growth, and yield in mixed northern conifers: 40-year results from the Penobscot Experimental Forest. Canadian Journal of Forest Research 33, 2116–2128.
Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64, 583–639.
Stein, M.L., 2007. Spatial variation of total column ozone on a global scale. Annals of Applied Statistics 1, 191–210.
Stein, M.L., 2008. A modeling approach for large spatial datasets. Journal of the Korean Statistical Society 37, 3–10.
Stein, M.L., Chi, Z., Welty, L.J., 2004. Approximating likelihoods for large spatial datasets. Journal of the Royal Statistical Society Series B 66, 275–296.
Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112, 1982–1999.
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, http://dx.doi.org/10.1016/j.rse.2011.05.009.
Vecchia, A.V., 1988. Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical Society Series B 50, 297–312.
Xia, G., Gelfand, A.E., 2006. Stationary process approximation for the analysis of large spatial datasets. Technical Report, ISDS, Duke University.
Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications. Springer-Verlag, Berlin.