HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2015 June 29.

Published in final edited form as:

J Am Stat Assoc. 2011 ; 106(493): 31–48. doi:10.1198/jasa.2011.ap09653.

A Hierarchical Model for Quantifying Forest Variables Over Large Heterogeneous Landscapes With Uncertain Forest Areas

Andrew O. Finley [Assistant Professor],
Departments of Forestry and Geography, Michigan State University, East Lansing, MI 48824 (finleya@msu.edu)

Sudipto Banerjee [Associate Professor of Biostatistics], and
School of Public Health, University of Minnesota, Minneapolis, MN 55455 (baner009@umn.edu)

David W. MacFarlane [Associate Professor of Forest Measurements and Modeling]
Departments of Forestry and Geography, Michigan State University, East Lansing, MI 48824 (macfar24@msu.edu)

Abstract

We are interested in predicting one or more continuous forest variables (e.g., biomass, volume, age) at a fine resolution (e.g., pixel level) across a specified domain. Given a definition of forest/nonforest, this prediction is typically a two-step process. The first step predicts which locations are forested. The second step predicts the value of the variable for only those forested locations. Rarely is the forest/nonforest status predicted without error. However, the uncertainty in this prediction is typically not propagated through to the subsequent prediction of the forest variable of interest. Failure to acknowledge this error can result in biased estimates of forest variable totals within a domain. In response to this problem, we offer a modeling framework that will allow propagation of this uncertainty. Here we envision two latent processes generating the data. The first is a continuous spatial process while the second is a binary spatial process. The continuous spatial process controls the spatial association structure of the forest variable of interest, while the binary process indicates presence of a possible nonzero value for the forest variable at a given location. The proposed models are applied to georeferenced National Forest Inventory (NFI) data and spatially coinciding remotely sensed predictor variables. Due to the large number of observed locations in this dataset we seek dimension reduction not just in the likelihood, but also for unobserved stochastic processes. We demonstrate how a low-rank *predictive process* can be adapted to our setting and reduce the dimensionality of the data and ease the computational burden.

Keywords

Bayesian inference; Biomass; Forestry; Markov chain Monte Carlo; Multivariate spatial process; Spatial models; Spatial predictive process

1. INTRODUCTION

Spatial prediction of forest variables is critical to many important contemporary global, regional, and local-scale decisions, including assessments of current carbon stock and flux,

biofeedstock for emerging bioeconomies, and impact of deforestation in both developed and developing nations. National forest inventory (NFI) programs often offer the data needed to support these assessments. Because complete enumerative inventories are prohibitively expensive, these programs generate coarse spatial resolution estimates of forest variables using samples from the population of interest and corresponding design-based estimators.

Although these estimators offer asymptotically unbiased estimates they are unable to produce prediction of forest variables at sufficiently fine spatial scales to meet information needs. Further, despite some advances in model-assisted methods (see, e.g., Opsomer et al. 2007), design-based estimators are not well suited to incorporating ancillary variables or complex spatial dependence structures to improve the accuracy and precision of parameter estimates and/or prediction. Therefore, model-based approaches to mapping are attracting greater interest. These initiatives typically entail statistical models that relate the forest variable of interest measured at a forest inventory plot to spatially coinciding predictors such as soil, climatic, topographic, and remotely sensed variables. These models are then used to produce digital data layers of small-area spatially explicit prediction across large domains, which ultimately support the end-user analyses described above. The common approach to mapping a forest variable is to discretize the predicted probability of forest occupancy at a location (e.g., if predicted probability >0.5 then set as forested, otherwise nonforested) and subsequently predict the forest variable of interest at only forested locations. This approach, however, does not propagate the uncertainty in the forest/nonforest prediction through to the predictive distribution of the forest variable of interest. It is unreasonable to assume the uncertainty due to forest/nonforest misclassification error is small. For instance, following the article by Diggle, Tawn, and Moyeed (1998) that detailed the use of spatial process models for non-Gaussian data, and similar work by Heagerty and Lele (1998) on a composite likelihood approach to binary spatial regression, Finley, Banerjee, and McRoberts (2008a) used a spatial logistic regression model and NFI data to predict forested areas. Their results showed substantial uncertainty in predicted probability of forest occupancy even with the use of informative remotely sensed predictor variables, dense array of forest/nonforest inventory plot locations, and strong spatial dependence.

Failure to acknowledge uncertainty in forest/nonforest classification can result in biased and perhaps misleading inference on the forest variable of interest. This is likely to become especially problematic when we wish to summarize the uncertainty in a forest variable over an area where forest occupancy must also be predicted. In response, we offer a modeling framework that will propagate this uncertainty, leverage the Euclidean proximity between national forest inventory (NFI) plot locations, and assess the potentially spatially varying impact of predictors on a forest variable of interest to improve the accuracy and precision of predictions at locations where we have observed these predictors, but not the inventory plots.

Here we envision two latent processes generating the data. The first is a continuous spatial process, while the second is a binary spatial process assumed a prior independent of the first. The continuous spatial process controls the spatial association structure of the forest variable of interest (e.g., Cressie 1993; Stein 1999), while a binary process indicates presence of a possible nonzero value for the forest variable at a given location. We advocate a hierarchical

model-based approach that allows for explicit quantification of forest area uncertainty as a component of the final variability in the estimate of the forest variable of interest. This model also allows for different sets of predictors to explain variability in the continuous and binary outcomes. This is important because the factors that give rise to the patterns of forest across a given landscape (e.g., anthropogenic deforestation) are likely very different from those that dictate patterns in the forest variable of interest. For example, the amount of forest biomass that could occur at a location is a function of microsite environmental variables (e.g., depth to water table, soil, aspect, elevation), forest structure and composition variables (e.g., age, density, species), regional variables (e.g., soil parent material, temperature extremes, precipitation), and disturbance history (e.g., hurricanes, drought, fire, timber harvesting). Further, the relationship between the outcomes (i.e., probability of forest occupancy and the forest variables of interest) and their respective sets of predictor variables are likely to vary over large domains. Therefore, when cast into a regression setting we find it attractive to allow the regression coefficients to vary by location, envisioning a spatial surface associated with each coefficient.

The hierarchical structure is attractive as it offers inference for the random spatial coefficients, and that too with non-Gaussian data, by delivering an entire posterior distribution at both observed and unobserved locations. This flexibility comes at an expense. When the number of observations is large, the Markov chain Monte Carlo (MCMC) methods (see, e.g., Robert and Casella 2005) used for estimation are prohibitively expensive. This modeling challenge has received much recent attention (see Section 3.1). Here we discuss how a low-rank spatial process can be adapted to model the regression coefficients and reduce the computational burden.

The proposed framework and novel application to forest inventory data presented here, represents the synthesis of modeling studies of continuous and categorical forest variables noted above. Specifically, by coupling a spatial logistic and Gaussian regression hierarchically, we are able to model the observed zero inflated continuous forest variable of interest. Further, within each level of the hierarchy we consider spatially varying regression models involving multivariate spatial processes (one for each regression coefficient) that we wish to model jointly. This approach is similar to Gelfand et al. (2003), but, unlike there, we allow each coefficient process to have its own spatial correlation structure. We achieve this using linear transformations of independent processes. This idea has been used elsewhere (e.g., Finley et al. 2008b; Finley, Banerjee, and McRoberts 2009a) to model multivariate continuous outcomes and spatially varying coefficients, but not completely unobserved coefficient processes nested within a hierarchical structure such as that encountered here.

The remainder of the article proceeds as follows. Section 2 provides an overview of the NFI data and study area used to motivate the models and methods presented in Sections 3. Section 4 presents a synthetic dataset analysis to explore properties of the proposed models. The full analysis of these NFI data is then offered in Section 5. Finally, Section 6 concludes the article with a brief summary and description of future direction and work.

2. DATA

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service conducts the NFI of the USA. The program has established field plot centers in permanent locations using a sampling design that produces an equal probability sample (Bechtold and Patterson 2005). The sampling design is based on a tessellation of the USA into approximately 2400 ha hexagons and features a permanent plot at a randomly selected location within each hexagon. The state of Michigan, in which the study area is located, provided additional funding to triple the sampling intensity to approximately one plot per 800 ha. Each plot consists of four 7.32 m radius circular subplots for a total area of 672 m^2. The subplots are configured as a central subplot and three peripheral subplots with centers located at 36.58 m and azimuths of 0°, 120°, and 240° from the center of the central subplot.

In general, locations of forested or previously forested plots are determined using Global Positioning System (GPS) receivers, whereas locations of nonforested plots are verified using aerial imagery and digitization methods. Field crews measure all trees with diameter at breast height (dbh) (1.37 m) of 12.7 cm or greater on forested plot. For each forested plot, total tree biomass is calculated with established equations that use tree-level variables including species, dbh, and total height. For our modeling purposes this value is expressed as metric tons of forest biomass per ha and associated with the coordinates of the center subplot. Subfigure (a) in Figure 1 illustrates the georeferenced forest inventory data consisting of 8774 forested FIA plots measured between 1999 and 2006 that meet our inclusion criterion. Here too, Subfigure (b) and (c) offer interpolated surfaces of forest/nonforest and forest biomass over the inventory plots.

The study area is the 107,011 km^2 lower peninsula Michigan. Approximately, 42,630 km^2, or 40%, of the this study area is forested. Precipitation and temperature extremes are well-known factors in determining spatial patterns of forest species composition and productivity in the Great Lakes states (Albert 1995). Therefore, we obtained raster data layers of mean annual precipitation (PRECIP), temperature minimum (TMIN), and temperature maximum (TMAX) over the period 1971–2000. These data were generated by the PRISM climate mapping project (Daly et al. 2000). In addition to these proxies for biological productivity we also use Normalized Difference Vegetation Index (NDVI) which is an index derived from Landsat satellite imagery which has been shown to be useful for predicting forest biomass (Dong et al. 2003). NDVI was calculated using a July 2003 mosaic of Landsat ETM + imagery. These data layers serve as predictor variables for the outcome variable metric tons of forest biomass per ha. The Landsat imagery was also used to derive tasseled cap components of brightness (TC1), greenness (TC2), and wetness (TC3) that are predictor variables used to distinguish forest from nonforest (Huang et al. 2002). The PRISM data layers were originally generated at ~0.8 × 0.8 km pixel resolution, but were resampled to match the 30 × 30 m resolution of the NDVI and tasseled cap data layers. We make this simplifying assumption because the disparity between the data layer resolution is small compared to the distance between FIA plot observations. Finally, all predictor data layers were reprojected to share the projection of the georeferenced forest inventory plots (Figure 2). Here we see strong spatial dependence within and among the predictor variables. For instance, thare are long north to south gradients in mean precipitation and temperature

extremes and gradients in temperature minimum that are perpendicular to the shorelines. The variables derived from the satellite imagery also show dependence but at a smaller spatial range compared to the climate variables. There is unquantified error in both the Landsat derived variables (e.g., sensor instrument and georegistration error) and PRISM climate data layers. For the current study, we assume these errors are small. This is a potentially naïve assumption, but one that is made by most who use these data for similar analyses.

3. MODELS

Let us consider the underlying mechanism generating the data. Assume we observe an outcome at a set of locations, say $\mathscr{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. We envision two latent processes generating these data. For a generic site **s**, the first process, **w**(**s**), is a continuous multivariate spatial process that accounts for *spatially varying regression coefficients* and incorporates spatial dependence in the outcome, as is customary in geostatistics (Cressie 1993). The second is a binary spatial process *z*(**s**) which is marginally independent of **w**(**s**) and indicates [i.e., *z*(**s**) = 1] presence of "measurable" biomass at a location. The observed process *y*(**s**) is now modeled using spatially varying regression coefficients

$$y(\mathbf{s}) = \left\{\mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s})'\mathbf{w}(\mathbf{s}) + \epsilon_1(\mathbf{s})\right\} z(\mathbf{s}) + \{1 - z(\mathbf{s})\}\,\epsilon_2(\mathbf{s}), \quad (1)$$

where $\epsilon_1(\mathbf{s}) \overset{iid}{\sim} \mathrm{N}\left(0, \tau_1^2\right)$ is the usual geostatistical nugget (measurement error or microscale variation) and $\epsilon_2(\mathbf{s}) \overset{iid}{\sim} \mathrm{N}\left(0, \tau_2^2\right)$ provides some additional noise to the structural zeros (see below). In (1), **x**(**s**) is the set of *p* regressors, $\tilde{\mathbf{x}}(\mathbf{s})$ includes those predictors from **x**(**s**) whose regression coefficients are posited to vary spatially. The process **w**(**s**) ~ MVGP(**0**, $\mathbf{C}_{\mathbf{w}}(\mathbf{s}_1, \mathbf{s}_2)$) denotes a *multivariate* Gaussian process completely characterized by their mean and a cross-covariance (matrix) function, $\mathbf{C}_{\mathbf{w}}(\mathbf{s}_1, \mathbf{s}_2; \boldsymbol{\theta}) = \mathrm{cov}\{\mathbf{w}(\mathbf{s}_1), \mathbf{w}(\mathbf{s}_2)\}$.

For locations where *z*(**s**) is known to be zero, *y*(**s**) is zero as well. Hence, τ_2^2 is not identifiable from the data; therefore, we fix it at some small value, for example, $\tau_2^2 = 0.0001$ to reflect a high-precision value around zero for unmeasurable *y*(**s**). Without the additional noise, ϵ_2(**s**), we encounter a degenerate likelihood that would preclude estimation. This avoids a degeneracy in the likelihood and renders stable estimation properties. Similar modifications to avoid degeneracy have appeared elsewhere in the analysis of challenging datasets (e.g., Higdon et al. 2008).

Cross-covariance functions are not routine to specify since they demand that for any number of locations and any choice of these locations the resulting covariance matrix for the associated process realizations be positive definite. Several different approaches ranging from moving averages to process convolutions have been explored; see, for example, Ver Hoef and Barry (1998), Higdon, Swall, and Kern (1999), Higdon, Lee, and Holloman (2003), Fuentes (2002), Gaspari and Cohn (1999), and Majumdar and Gelfand (2007). Recently, Gneiting, Kleiber, and Schlather (2009) proposed a multivariate Matérn correlation function as an alternative to defining covariances based on the LMC. An

attractive, easily interpretable and flexible approach develops versions of the Linear Model of Core-gionalization (LMC) as in, for example, Grzebyk and Wackernagel (1994), Wackernagel (2006), Schmidt and Gelfand (2003), or Gelfand et al. (2004). Reich and Fuentes (2007) offer a Bayesian nonparametric adaptation.

The LMC builds valid multivariate process models with a (possibly) space-varying linear transformation $\mathbf{w}(\mathbf{s}) = \mathbf{L}(\mathbf{s})\mathbf{e}(\mathbf{s})$, where $\mathbf{L}(\mathbf{s})$ is a nonsingular $p \times p$ matrix, $\mathbf{e}(\mathbf{s})$ $(e_1(\mathbf{s}), \ldots, e_p(\mathbf{s}))'$ and each $e_i(\mathbf{s})$ is an independent spatial=process with unit variance and correlation function $\rho_i(\mathbf{s}_1, \mathbf{s}_2; \theta_i)$. Thus, $\mathbf{e}(\mathbf{s})$ has a diagonal cross-covariance matrix $\mathbf{C_e}(\mathbf{s}_1, \mathbf{s}_2)$ with ith diagonal element as $\rho_i(\mathbf{s}_1, \mathbf{s}_2; \theta_i)$ yielding a valid nonstationary cross-covariance $\mathbf{C_w}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{L}(\mathbf{s}_1)\mathbf{C_e}(\mathbf{s}_1, \mathbf{s}_2)\mathbf{L}(\mathbf{s}_2)'$ for $\mathbf{w}(\mathbf{s})$. A particularly flexible choice for the correlation function is the Matérn, which allows control of spatial decay, ϕ_1, and smoothness, ϕ_2 (Stein 1999) and is given by

$$\begin{aligned}\rho(\mathbf{s}_1, \mathbf{s}_2; \boldsymbol{\phi}) = \; & \frac{1}{2^{\phi_2 - 1}\Gamma(\phi_2)}(\| \mathbf{s}_1 - \mathbf{s}_2 \| \phi_1)^{\phi_2} \\ & \times \mathcal{K}_{\phi_2}(\| \mathbf{s}_1 - \mathbf{s}_2 \| ; \phi_1); \qquad \phi_1 > 0, \phi_2 > 0,\end{aligned} \tag{2}$$

In general, $\mathbf{w}(\mathbf{s})$ is nonstationary even when $\mathbf{e}(\mathbf{s})$ is stationary. When $\mathbf{L}(\mathbf{s}) = \mathbf{L}$ is constant, $\mathbf{w}(\mathbf{s})$ inherits stationarity from $\mathbf{e}(\mathbf{s})$; $\mathbf{C_w}(\mathbf{s}_1 - \mathbf{s}_2) = \mathbf{L}\mathbf{C_e}(\mathbf{s}_1 - \mathbf{s}_2)\mathbf{L}'$. Since $\mathbf{C_w}(\mathbf{s}, \mathbf{s})$ $\mathbf{L}(\mathbf{s})\mathbf{L}(\mathbf{s})'$, without loss of generality one can assume that $\mathbf{L}(\mathbf{s}) = \mathbf{C}_{\mathbf{w}}^{1/2}(\mathbf{s}, \mathbf{s})$ is a lower-triangular square-root; the one-to-one correspondence between the elements of $\mathbf{L}(\mathbf{s})$ and $\mathbf{C_w}(\mathbf{s}, \mathbf{s})$ is well known (see, e.g., Harville 1997, p. 229). Thus, $\mathbf{L}(\mathbf{s})$ determines the association between the elements of $\mathbf{w}(\mathbf{s})$ within $\mathbf{s}$. Choices for modeling $\mathbf{L}(\mathbf{s})$ include an inverse spatial-Wishart process for $\mathbf{L}(\mathbf{s})\mathbf{L}(\mathbf{s})'$ (Gelfand et al. 2004) or element-wise modeling with Gaussian and log-Gaussian processes. Stationarity implies $\mathbf{L}(\mathbf{s}) = \mathbf{L}$ and that $\mathbf{C_w}(\mathbf{0}) = \mathbf{L}\mathbf{L}'$. Here we could either assign a prior, for example, inverse Wishart, to $\mathbf{L}\mathbf{L}'$ or could further parameterize it in terms of eigenvalues and the Givens angles which are themselves assigned hyperpriors (Daniels and Kass 1999).

Assuming $\tilde{\mathbf{x}}(\mathbf{s}) = \mathbf{x}(\mathbf{s})$ in the subsequent development, let $\boldsymbol{\Sigma}_{\mathbf{w}}$ denote the $np \times np$ matrix partitioned into $p \times p$ blocks with $\mathbf{C_w}(\mathbf{s}_i, \mathbf{s}_j)$ forming the (i, j)th block. Now, we write the hierarchical model:

$$y(\mathbf{s}_i) | \boldsymbol{\beta}, \mathbf{w}(\mathbf{s_i}), \mathbf{z}(\mathbf{s_i}) \overset{\text{iid}}{\sim} \mathrm{N}\left(\left\{\mathbf{x}(\mathbf{s_i})'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s_i})'\mathbf{w}(\mathbf{s_i})\right\}\mathbf{z}(\mathbf{s_i}), \mathbf{z}(\mathbf{s_i})\tau_1^2 + (\mathbf{1} - \mathbf{z}(\mathbf{s_i}))\tau_2^2\right), \quad \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n};$$

$$\mathbf{w} | \mathbf{L}, \phi \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}}), \qquad \text{where} \quad \mathbf{w} = \left(\mathbf{w}(\mathbf{s_1})', \ldots, \mathbf{w}(\mathbf{s_n})'\right)';$$

$$\begin{aligned} z(\mathbf{s}_i) | \boldsymbol{\eta}, \mathbf{u}(\mathbf{s_i}) \tilde{} \; & Ber(\pi(\mathbf{s}_i)); \\ logit(\pi(\mathbf{s}_i)) \; & = \mathbf{v}(\mathbf{s}_i)'\boldsymbol{\eta} + \tilde{\mathbf{v}}(\mathbf{s_i})'\mathbf{u}(\mathbf{s_i}), \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}; \end{aligned}$$

$$\mathbf{u}|\mathbf{L}_z, \phi_z \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{u}), \qquad \text{where} \quad \mathbf{u} = \left(\mathbf{u}(\mathbf{s}_1)', \ldots, \mathbf{u}(\mathbf{s}_n)'\right)';$$

$$\boldsymbol{\beta} \sim p(\boldsymbol{\beta}); \qquad \mathbf{L} \sim p(\mathbf{L});$$

$$\phi \sim p(\phi); \qquad \tau_1^2 \sim p\left(\tau_1^2\right); \qquad \tau_2^2 \sim p\left(\tau_2^2\right);$$

$$\boldsymbol{\eta} \sim p(\boldsymbol{\eta}); \qquad \mathbf{L}_\mathbf{z} \sim p(\mathbf{L}_\mathbf{z}); \qquad \phi_\mathbf{z} \sim p(\phi_\mathbf{z}).$$

The above framework is very flexible. It accommodates two multivariate spatial processes—**w**(**s**) for the biomass measure and an analogous process **u**(**s**) ~ MVGP(0, $\mathbf{C}_z(\cdot;\mathbf{L}_z, \phi_z)$) [with $\mathbf{L}_z$ and ϕ_z defined analogous to **L** and ϕ for **w**(**s**)] which controls spatial association between locations with structured zeros. If this is undesirable, we can simply fix **u**(**s**) = 0. The error-process ϵ_2(**s**) may or may not have a strong physical interpretation, but it is crucial for the first stage of the above hierarchical model to be conditional normals.

Let $\boldsymbol{\theta} = \left\{\tau_1^2, \tau_2^2, \mathbf{L}, \phi, \mathbf{L}_\mathbf{z}, \phi_\mathbf{z}\right\}$. For our subsequent development, we find it convenient to further define $\boldsymbol{\theta}_1 = \left\{\tau_1^2, \tau_2^2\right\}$, $\boldsymbol{\theta}_2$ = {**L**, ϕ}, and $\boldsymbol{\theta}_3$ = {$\mathbf{L}_z$, ϕ_z}. The joint posterior distribution $p(\boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\eta}, \mathbf{u}, \boldsymbol{\theta}|\mathbf{y})$ is proportional to

$$\begin{aligned} p(\boldsymbol{\theta}) \times \quad & \mathrm{N}(\mathbf{u}|\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{u}) \times p(\boldsymbol{\eta}) \\ & \times \prod_{i=1}^{n} Ber(z(\mathbf{s}_i)|\pi(\mathbf{s}_i)) \times \mathrm{N}(\mathbf{w}|\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{w}) \times p(\boldsymbol{\beta}) \\ & \times \prod_{i=1}^{n} \mathrm{N}\left(y(\mathbf{s}_i) \mid \left\{\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s}_\mathbf{i})'\mathbf{w}(\mathbf{s}_\mathbf{i})\right\} z(\mathbf{s}_i), z(\mathbf{s}_i)\tau_1^2 + (1 - z(\mathbf{s}_i))\tau_2^2\right), \end{aligned} \tag{3}$$

where **y** = (y(**s**$_1$), …, y(**s**$_n$))′ and **z** analogously denotes the observed binary data vector. Also,

$$\begin{aligned} Ber(z(\mathbf{s}_i)|\pi(\mathbf{s}_i)) = & \left(\frac{exp\left(\mathbf{v}(\mathbf{s}_i)'\boldsymbol{\eta} + \tilde{\mathbf{v}}(\mathbf{s}_\mathbf{i})'\mathbf{u}(\mathbf{s}_\mathbf{i})\right)}{1 + exp\left(\mathbf{v}(\mathbf{s}_i)'\boldsymbol{\eta} + \tilde{\mathbf{v}}(\mathbf{s}_\mathbf{i})'\mathbf{u}(\mathbf{s}_\mathbf{i})\right)}\right)^{z(\mathbf{s}_i)} \\ & \times \left(\frac{1}{1 + exp\left(\mathbf{v}(\mathbf{s}_i)'\boldsymbol{\eta} + \tilde{\mathbf{v}}(\mathbf{s}_\mathbf{i})'\mathbf{u}(\mathbf{s}_\mathbf{i})\right)}\right)^{1 - z(\mathbf{s}_i)}, \end{aligned}$$

where **v**(**s**) and $\tilde{\mathbf{v}}(\mathbf{s})$ are defined analogously to **x**(**s**$_i$) and $\tilde{\mathbf{x}}(\mathbf{s}_i)$. An MCMC algorithm for estimating the above model can update $\boldsymbol{\beta}$ and **w** from Gibbs updates. With a N($\boldsymbol{\mu}_\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_\boldsymbol{\beta}$) prior for $\boldsymbol{\beta}$, we update

$$\boldsymbol{\beta}|\cdot \sim \mathrm{N}\left(\left[\boldsymbol{\Sigma}_\boldsymbol{\beta}^{-1} + \mathbf{X}'\mathbf{Z}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{Z}\mathbf{X}\right]^{-1}\left\{\boldsymbol{\Sigma}_\boldsymbol{\beta}^{-1}\boldsymbol{\mu}_\boldsymbol{\beta} + \mathbf{X}'\mathbf{Z}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\left(\mathbf{y} - \mathbf{Z}\tilde{\mathbf{X}}\mathbf{w}\right)\right\}, \left[\boldsymbol{\Sigma}_\boldsymbol{\beta}^{-1} + \mathbf{X}'\mathbf{Z}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{Z}\mathbf{X}\right]^{-1}\right), \tag{4}$$

where $\mathbf{X} = (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$, $\tilde{\mathbf{X}}$ is an $n \times np$ block-diagonal matrix with $\tilde{\mathbf{x}}(\mathbf{s}_i)'$ as the ith diagonal, $\mathbf{Z}$ and $\boldsymbol{\Sigma}_y$ are $n \times n$ diagonal matrices with $z(\mathbf{s}_i)$ and $z(\mathbf{s}_i)\tau_1^2 + (1 - z(\mathbf{s}_i))\tau_2^2$ as their ith diagonal elements, respectively. If we use a flat prior for $\boldsymbol{\beta}$, we set $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} = O$, the null matrix. The spatial effects $\mathbf{w}$ can also be updated from a closed-form kernel:

$$\mathbf{w} | \cdot \sim \mathrm{N}\left(\left[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \tilde{\mathbf{X}}'\mathbf{Z}'\boldsymbol{\Sigma}_y^{-1}\mathbf{Z}\tilde{\mathbf{X}}\right]^{-1}\left\{\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\mu}_{\mathbf{w}} + \tilde{\mathbf{X}}'\mathbf{Z}'\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mathbf{Z}\mathbf{X}\boldsymbol{\beta})\right\}, \left[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \tilde{\mathbf{X}}'\mathbf{Z}'\boldsymbol{\Sigma}_y^{-1}\mathbf{Z}\tilde{\mathbf{X}}\right]^{-1}\right), \quad (5)$$

where $\boldsymbol{\Sigma}_{\mathbf{w}}$ is an $np \times np$ matrix partitioned into $p \times p$ blocks with $\mathbf{C}_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j$ forming the (i, j)th block, and $\boldsymbol{\mu}_{\mathbf{w}}$ is the mean of the Gaussian process $\mathbf{w}(\mathbf{s})$, taken to be $\mathbf{0}$ here. All the remaining parameters, including the latent process realization $\mathbf{u}$ will require metropolis steps.

After convergence, we obtain posterior samples $\left\{\boldsymbol{\beta}^{(k)}, \mathbf{w}^{(k)}, \boldsymbol{\eta}^{(k)}, \mathbf{u}^{(k)}, \boldsymbol{\theta}^{(k)}\right\}_{k=1}^{K}$. We can now sample from the posterior predictive distribution of the dependent variable at an arbitrary location. We consider two cases: to evaluate the fit of the model we would like to obtain the predictive distribution at observed locations, that is, $\mathbf{s}_i \in \mathscr{S}$, by evaluating

$$\begin{aligned} p(\tilde{y}(\mathbf{s}_i) | \mathbf{y}, \mathbf{z}) = & \int \mathrm{N}\left(\tilde{y}(\mathbf{s}_i) \mid \left\{\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s}_i)'\mathbf{w}(\mathbf{s}_i)\right\} z(\mathbf{s}_i), z(\mathbf{s}_i)\tau_1^2 + (1 - z(\mathbf{s}_i))\tau_2^2\right) \\ & \times p(\boldsymbol{\beta}, \mathbf{w}, \tau_1^2, \tau_2^2 | \mathbf{y}, \mathbf{z})\, d\boldsymbol{\beta} \quad d\mathbf{w} \quad d\boldsymbol{\eta} \quad d\mathbf{u} \quad d\boldsymbol{\theta}. \end{aligned} \quad (6)$$

For each $\left\{\boldsymbol{\beta}^{(k)}, \mathbf{w}(\mathbf{s}_i)^{(k)}, \tau_1^{2(k)}, \tau_2^{2(k)}\right\}$, we draw

$$\tilde{y}(\mathbf{s}_i)^{(k)} \sim \mathrm{N}\left(\left\{\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}^{(k)} + \tilde{\mathbf{x}}(\mathbf{s}_i)'\mathbf{w}(\mathbf{s}_i)^{(k)}\right\} z(\mathbf{s}_i), z(\mathbf{s}_i)\tau_1^{2(k)} + (1 - z(\mathbf{s}_i))\tau_2^{2(k)}\right).$$

The resulting $\left\{\tilde{y}(\mathbf{s}_i)^{(k)}\right\}_{k=1}^{K}$ are samples from (6) and can be used to summarize posterior predictive distributions of the fitted values. Turning to prediction at a location $\mathbf{s}_0 \notin \mathscr{S}$, we seek

$$\begin{aligned} & p(\tilde{y}(\mathbf{s}_0), \tilde{z}(\mathbf{s}_0) | \mathbf{y}, \mathbf{z}) \\ & = \int p(\tilde{y}(\mathbf{s}_0) | \tilde{z}(\mathbf{s}_0), \boldsymbol{\beta}, \mathbf{w}(\mathbf{s}_0), \tau_1^2, \tau_2^2) \\ & \quad \times p(\tilde{z}(\mathbf{s}_0) | \boldsymbol{\eta}, \mathbf{u}(\mathbf{s}_0)) \\ & \quad \times p(\boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\eta}, \mathbf{u}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{z})\, d\boldsymbol{\beta} \quad d\mathbf{w} \quad d\boldsymbol{\eta} \quad d\mathbf{u} \quad d\boldsymbol{\theta}. \end{aligned} \quad (7)$$

We draw samples from (7) using posterior predictive composition sampling. For each $\{\boldsymbol{\eta}^{(k)}, \mathbf{u}(\mathbf{s}_0)^{(k)}$ we sample $\tilde{z}(\mathbf{s}_0)^{(k)} \sim Ber\left(\pi^{(k)}(\mathbf{s}_0)\right)$, where

$\pi^{(k)}(\mathbf{s}_0) = exp\left(\mathbf{v}(\mathbf{s}_0)'\boldsymbol{\eta}^{(k)} + \tilde{\mathbf{v}}(\mathbf{s}_i)' \times \mathbf{u}(\mathbf{s}_0)^{(k)}\right) / \left(1 + exp\left(\mathbf{v}(\mathbf{s}_0)'\boldsymbol{\eta}^{(k)} + \tilde{\mathbf{v}}(\mathbf{s}_i)'\mathbf{u}(\mathbf{s}_0)^{(k)}\right)\right)$.

Next, for each $\left\{z(\mathbf{s}_0)^{(k)}, \boldsymbol{\beta}^{(k)}, \mathbf{w}(\mathbf{s}_0)^{(k)}, \tau_1^{2(k)}, \tau_2^{2(k)}\right\}$ we draw

$$\tilde{y}(\mathbf{s}_0)^{(k)} \sim \mathrm{N}\left(\left\{\mathbf{x}(\mathbf{s}_0)'\boldsymbol{\beta}^{(k)} + \tilde{\mathbf{x}}(\mathbf{s}_0)'\mathbf{w}(\mathbf{s}_0)^{(k)}\right\} \tilde{z}(\mathbf{s}_0)^{(k)}, \tilde{z}(\mathbf{s}_0)^{(k)}\tau_1^{2(k)} + \left(1 - \tilde{z}(\mathbf{s}_0)^{(k)}\right)\tau_2^{2(k)}\right).$$

For this scheme we require $\mathbf{u}(\mathbf{s}_0)^{(k)}$ and $\mathbf{w}(\mathbf{s}_0)^{(k)}$. When $\mathbf{s}_0 \notin \mathscr{S}$ is a new location, we need an additional step: we draw

$$\mathbf{w}(\mathbf{s}_0)^{(k)} \sim \mathrm{N}\left(\mathbf{C}\left(\mathbf{s}_0;\boldsymbol{\theta}_2^{(k)}\right)'\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{w}^{(k)}, \left(\mathbf{L}^{(k)}\mathbf{L}'^{(k)} - \mathbf{C}\left(\mathbf{s}_0;\boldsymbol{\theta}_2^{(k)}\right)'\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}\left(\mathbf{s}_0;\boldsymbol{\theta}_2^{(k)}\right)\right)\right),$$

$$\mathbf{u}(\mathbf{s}_0)^{(k)} \sim \mathrm{N}\left(\mathbf{C}_z\left(\mathbf{s}_0;\boldsymbol{\theta}_3^{(k)}\right)'\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{u}^{(k)}, \left(\mathbf{L}_z^{(k)}\mathbf{L}_z'^{(k)} - \mathbf{C}_z\left(\mathbf{s}_0;\boldsymbol{\theta}_3^{(k)}\right)'\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}\mathbf{C}_z\left(\mathbf{s}_0;\boldsymbol{\theta}_3^{(k)}\right)\right)\right),$$

where $\mathbf{C}\left(\mathbf{s}_0;\boldsymbol{\theta}_2^{(k)}\right)$ is an $np \times p$ matrix partitioned into an $n \times 1$ block matrix whose ith block is the $p \times p$ matrix $\mathbf{C}_{\mathbf{w}}(\mathbf{s}_0, \mathbf{s}_i)$ while $\mathbf{C}_z(\mathbf{s}_0; \boldsymbol{\theta}_3)$ is defined analogously using $\mathbf{C}_z(\mathbf{s}_0, \mathbf{s}_i)$.

3.1 Predictive Process Models

For a large number of spatial locations, fitting customary geostatistical models becomes prohibitive with necessary matrix factorizations of cubic order complexities. Without further specifications, estimating (3) will involve matrix computations involving the $np \times np$ matrix $\boldsymbol{\Sigma}_{\mathbf{w}}$. Such computations invoke linear solvers or Cholesky decompositions of complexity $O(n^3p^3)$, once every MCMC iteration, to produce estimates of $\boldsymbol{\theta}$. With large n, this is computationally infeasible.

Modeling large spatial datasets have received much attention in the recent past. One popular approach is to consider a smaller set of locations, or "knots," say $\mathscr{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_{n^*}^*\}$, where the number of knots, n^*, is much smaller than the number of observed sites, and to express the spatial process realizations over S in terms of its realizations over the smaller set of knots. The underlying rationale is that the spatial information available from the n sites could, in fact, be effectively captured through a smaller, but representative, set of locations —the knots. These are often referred to as low-rank or reduced-rank spatial models and have been explored in different contexts.

There are several variants of low-rank spatial models available to modelers (e.g., Wahba 1990; Higdon 2002; Rasmussen and Williams 2006; Stein 2007, 2008; Banerjee et al. 2008; Crainiceanu, Diggle, and Rowlingson 2008; Cressie and Johannesson 2008). We seek a dimension-reducing process that will not only integrate seamlessly into our hierarchical framework, but also will easily adapt to multivariate processes. The multivariate processes are needed because we posit associations among the spatially varying coefficients. In addition, low-rank models often lead to oversmoothing that produce biased estimates for the geostatistical nugget and the spatial variance components. We seek to rectify or offset this bias.

Based upon the above considerations, we opt for a class of *predictive process* models that emerge from optimal projections of the original spatial process in a lower-dimensional subspace. We refer to the $\mathbf{w}(\mathbf{s})$ as the *parent process* and consider its realizations over an arbitrary set of locations, say $\mathscr{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_{n^*}^*\}$, called the "knots" in the spatial domain.

Letting $\mathbf{w}^*=\{\mathbf{w}(\mathbf{s}_i^*)\}_{i=1}^{n^*}\tilde{}MVN(\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2))$ denote the realizations of **w(s)** over $\mathscr{S}^*$, with $\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)$ being the corresponding $n^*p \times n^*p$ blocked covariance matrix whose (i, j)th block is given by the $p \times p$ cross-covariance matrix $\mathbf{C}_{\mathbf{w}}\left(\mathbf{s}_i^*,\mathbf{s}_j^*\right)=cov\left\{\mathbf{w}(\mathbf{s}_i^*),\mathbf{w}\left(\mathbf{s}_j^*\right)\right\}$, Banerjee et al. (2008) propose a *predictive process* at a generic site **s** derived from $\tilde{\mathbf{w}}(\mathbf{s})$ as

$$\tilde{\mathbf{w}}(\mathbf{s})=E\left[\mathbf{w}(\mathbf{s})\,|\mathbf{w}^*\right]=\mathbf{C}(\mathbf{s};\boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}\mathbf{w}^*, \quad (8)$$

where $\mathbf{C}(\mathbf{s};\boldsymbol{\theta}_2)=\left\{\mathbf{C}_{\mathbf{w}}\left(\mathbf{s},\mathbf{s}_j^*;\boldsymbol{\theta}_2\right)\right\}_{j=1}^{n^*}$ is a $p \times n^*p$ matrix comprising the cross-covariances between the parent process **w(s)** at a generic site **s** and the realizations of the parent process over the collection of knots. The process $\tilde{\mathbf{w}}(\mathbf{s})$ is called the *predictive process* (Banerjee et al. 2008) derived from the parent process **w(s)**. [More precisely, one should write $\tilde{\mathbf{w}}_{\mathscr{S}^*}(\mathbf{s})$, but this dependence is implicit.] This process has well-defined realizations (i.e., yield nonsingular joint distributions) over sets with at most n^* locations.

The predictive process is smoother than the original process and systematically underestimates the variance of the parent process w**(s)** at any location **s** (Finley et al. 2009b). However, a *bias-adjusted predictive process* can be formed as

$$\begin{aligned}\tilde{\mathbf{w}}_{\tilde{\epsilon}}(\mathbf{s}) &= \tilde{\mathbf{w}}(\mathbf{s})+\tilde{\epsilon}(\mathbf{s});\\ \tilde{\epsilon}(\mathbf{s}) &\overset{ind}{\sim} \mathrm{N}\left(0,\left\{\mathbf{L}\mathbf{L}'-\mathbf{C}(\mathbf{s};\boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}\mathbf{C}(\mathbf{s};\boldsymbol{\theta}_2)\right\}\right).\end{aligned} \quad (9)$$

Given parameters in $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, we also make use of the following matrices: $\boldsymbol{\Sigma}_{\tilde{\epsilon}}$ is the $n \times n$ diagonal matrix with ith diagonal element given by $\{\mathbf{L}\mathbf{L}' - \mathbf{C}(\mathbf{s}_i; \boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}\mathbf{C}(\mathbf{s}_i; \boldsymbol{\theta}_2)\}$, $\boldsymbol{\Sigma}_{\epsilon}=\tau_1^2\boldsymbol{Z}+\tau_2^2(\mathbf{I}-\boldsymbol{Z})$ (an $n \times n$ diagonal matrix) and $\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}=\tilde{\mathbf{X}}\boldsymbol{\Sigma}_{\tilde{\epsilon}}\tilde{\mathbf{X}}'+\boldsymbol{\Sigma}_{\epsilon}$. Note that $\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}$ is again a diagonal matrix with the ith diagonal element being given by $\tilde{\mathbf{x}}(\mathbf{s}_i)'\left\{\mathbf{L}\mathbf{L}'-\mathbf{C}(\mathbf{s}_i;\boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}\mathbf{C}(\mathbf{s}_i;\boldsymbol{\theta}_2)\right\}\tilde{\mathbf{x}}(\mathbf{s}_i)+\tau_1^2 z(\mathbf{s}_i)+\tau_2^2(1-z(\mathbf{s}_i))$.

Replacing the parent processes **w(s)** and **u(s)** with their modified predictive process counterparts, $\tilde{\mathbf{w}}_{\tilde{\epsilon}}(\mathbf{s})$ and $\tilde{\mathbf{u}}_{\tilde{\epsilon}}(\mathbf{s})$ in our hierarchical model results in the following predictive process version of (3):

$$\begin{aligned}p(\boldsymbol{\theta})\times\ & \mathrm{N}(\mathbf{u}^*|\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{u}^*})\times\mathrm{N}(\tilde{\mathbf{u}}_{\tilde{\epsilon}}|\mathbf{F}_u(\boldsymbol{\theta}_3)\mathbf{u}^*,\boldsymbol{\Sigma}_{\tilde{\epsilon}_u})\\ &\times p(\boldsymbol{\eta})\times\prod_{i=1}^{n}Ber(z(\mathbf{s}_i)\,|\pi(\mathbf{s}_i))\\ &\times\mathrm{N}(\mathbf{w}^*|\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2))\times\mathrm{N}(\tilde{\mathbf{w}}_{\tilde{\epsilon}}|\mathbf{F}(\boldsymbol{\theta}_2)\mathbf{w}^*,\boldsymbol{\Sigma}_{\tilde{\epsilon}})\times p(\boldsymbol{\beta})\\ &\times\prod_{i=1}^{n}\mathrm{N}\left(y(\mathbf{s}_i)\,|\left\{\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}+\tilde{\mathbf{x}}(\mathbf{s}_i)'\tilde{\mathbf{w}}_{\tilde{\epsilon}}(\mathbf{s})\right\}z(\mathbf{s}_i),z(\mathbf{s}_i)\tau_1^2+(1-z(\mathbf{s}_i))\tau_2^2\right),\end{aligned} \quad (10)$$

where $\tilde{\mathbf{w}}_{\tilde{\epsilon}}=(\tilde{\mathbf{w}}(\mathbf{s}_1),\ldots,\tilde{\mathbf{w}}(\mathbf{s}_n))'$, $\mathbf{F}(\boldsymbol{\theta}_2)=\mathscr{C}(\boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}$ and $\mathscr{C}(\boldsymbol{\theta}_2)'$ is an $np \times n^*p$ block matrix whose (i, j)th block is $\mathbf{C}_{\mathbf{w}}\left(\mathbf{s}_i,\mathbf{s}_j^*;\phi\right)$. The definitions of $\mathbf{u}^*$, $\tilde{\mathbf{u}}_{\tilde{\epsilon}}$, $\mathbf{F}_u(\boldsymbol{\theta}_3)$ and $\boldsymbol{\Sigma}_{\tilde{\epsilon}_u}$ are analogous to those for the **w(s)** process. The bias-adjustment does not result in the

introduction of new parameters from what were present in the parent model, hence all model parameters are identifiable.

Estimation proceeds using Markov chain Monte Carlo (MCMC) employing a Gibbs sampler with Metropolis–Hastings steps (Gelman et al. 2004). Here it is often easier to use a marginalized likelihood after integrating out the spatial random effects **w*** and $\tilde{\mathbf{w}}_{\tilde{\epsilon}}$. The joint posterior is given by

$$p(\boldsymbol{\theta})\times \mathrm{N}(\mathbf{u}|\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{u}})\times p(\boldsymbol{\eta})\times\prod_{i=1}^{n} Ber(z(\mathbf{s}_i)|\boldsymbol{\pi}(\mathbf{s_i}))\times p(\boldsymbol{\beta})\times MVN\left(\mathbf{y}|\mathbf{Z}\mathbf{X}\boldsymbol{\beta},\mathscr{C}(\boldsymbol{\theta_2})'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta_2})^{-1}\mathscr{C}(\boldsymbol{\theta_2})+\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}\right). \quad (11)$$

We use a Gibbs update on $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|\cdot \sim \mathbf{N}\left(\left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}+\mathbf{X}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{X}\right]^{-1}\left\{\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}+\mathbf{X}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}\right\},\left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}+\mathbf{X}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{X}\right]^{-1}\right), \quad (12)$$

where $\boldsymbol{\Sigma}_y=\mathscr{C}(\boldsymbol{\theta}_2)'\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}\mathscr{C}(\boldsymbol{\theta}_2)+\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}$. we invoke the Sherman–Woodbury–Morrison (SWM) matrix identity (Henderson and Searle 1981) to write

$$\boldsymbol{\Sigma}_y^{-1}=\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}-\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}\mathscr{C}(\boldsymbol{\theta}_2)'\times\left[\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)+\mathscr{C}(\boldsymbol{\theta}_2)\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}\mathscr{C}(\boldsymbol{\theta}_2)'\right]^{-1}\mathscr{C}(\boldsymbol{\theta}_2)\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}. \quad (13)$$

The above computation involves $\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}$ which is diagonal and the inverse of the $n^*p \times n^*p$ matrix $\left[\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)+\mathscr{C}(\boldsymbol{\theta}_2)\times\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}\mathscr{C}(\boldsymbol{\theta}_2)'\right]$ which is cheaper than computing $\boldsymbol{\Sigma}_y^{-1}$.

Once the parameters in $\boldsymbol{\theta}$ have been sampled, we can recover the predictive process spatial effects as follows. For each post burn-in sample $\boldsymbol{\theta}^{(k)}$, we update $\mathbf{w}^{*(k)} \sim \mathrm{N}(\boldsymbol{\mu}_{\mathbf{w}^*|\cdot}, \boldsymbol{\Sigma}_{\mathbf{w}^*|\cdot})$:

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{w}^*|\cdot} &= \left[\boldsymbol{\Sigma}_{\mathbf{w}^*}(\boldsymbol{\theta}_2)^{-1}+\mathbf{F}(\boldsymbol{\theta}_2)'\tilde{\mathbf{X}}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}\boldsymbol{Z}\tilde{\mathbf{X}}\mathbf{F}(\boldsymbol{\theta}_2)\right]^{-1};\\ \boldsymbol{\mu}_{\mathbf{w}^*|\cdot} &= \boldsymbol{\Sigma}_{\mathbf{w}^*|\cdot}\mathbf{F}(\boldsymbol{\theta}_2)'\tilde{\mathbf{X}}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}^{-1}\left(\mathbf{y}-\mathbf{Z}\mathbf{X}\boldsymbol{\beta}^{(k)}\right).\end{aligned}$$

Subsequently, we sample the "adjustment error" term $\tilde{\epsilon}^{(k)}$ from $\mathrm{N}\left(\boldsymbol{\mu}_{\tilde{\epsilon}|\cdot}, \boldsymbol{\Sigma}_{\tilde{\epsilon}|\cdot}\right)$, where

$$\begin{aligned}\boldsymbol{\Sigma}_{\tilde{\epsilon}|\cdot} &= \left[\boldsymbol{\Sigma}_{\tilde{\epsilon}}\left(\boldsymbol{\theta}_2^{(k)}\right)^{-1}+\tilde{\mathbf{X}}'\boldsymbol{Z}'\boldsymbol{\Sigma}_{\epsilon}\left(\boldsymbol{\theta}_1^{(k)}\right)^{-1}\boldsymbol{Z}\tilde{\mathbf{X}}\right]^{-1};\\ \boldsymbol{\mu}_{\tilde{\epsilon}|\cdot} &= \boldsymbol{\Sigma}_{\tilde{\epsilon}|\cdot}\boldsymbol{\Sigma}_{\epsilon}^{-1}\left(\mathbf{y}-\mathbf{Z}\mathbf{X}\boldsymbol{\beta}^{(k)}-\boldsymbol{Z}\tilde{\mathbf{X}}\mathbf{F}\left(\boldsymbol{\theta}_2^{(k)}\right)\mathbf{w}^{*(k)}\right),\end{aligned}$$

and compute $\tilde{\mathbf{w}}_{\tilde{\epsilon}}^{*(k)}=\mathbf{F}\left(\boldsymbol{\theta}_2^{(k)}\right)\mathbf{w}^{*(k)}+\tilde{\epsilon}^{(k)}$. Finally, prediction and interpolation can be carried out in a manner analogous to that described in the preceding section but with the predictive process counterparts of **w**(**s**) and **u**(**s**) in (6) and (7).

Given the reliance on knots, we must decide how to chose the configuration of the knot array. With a fairly even distribution of data locations, one possibility is to select knots on a uniform grid overlaid on the domain. Alternatively, selection can be achieved through a

formal design-based approach based upon minimization of a spatially averaged predictive variance criterion (see, e.g., Diggle and Lophaven 2006). However, in general the locations are highly irregular, generating substantial areas of sparse observations where we wish to avoid placing knots, since they would be "wasted" and possibly lead to inflated predictive process variances and slower convergence. Here, space-covering designs (e.g., Royle and Nychka 1998) can yield a representative collection of knots that better cover the domain. Finley et al. (2009b) proposed an alternative algorithm that uses an objective function that minimizes average predictive variance to choose a set of knots from a dense grid of candidate locations. This approach, however, is too computationally demanding in the current setting. A simpler and computationally less demanding approach is to apply popular clustering algorithms such as k-means or more robust median-based *partitioning around medoids* algorithms (e.g., Kaufmann and Roseauw 1990). User-friendly implementations of these algorithms are available in R packages such as fields and cluster and have been used in spline-based low-rank kriging models (e.g., Lin et al. 2000; Ruppert, Wand, and Carroll 2003; Crainiceanu, Diggle, and Rowlingson 2008). For the current analysis, we experimented with these clustering algorithms and the infill designs suggested by Diggle and Lophaven (2006); however, neither provided changes in model parameter estimates or improvements in prediction. We note, however, this lack of improvement is likely data specific and these knot placement approaches should be explored for each new analysis.

4. ANALYSIS OF SYNTHETIC DATA

4.1 Implementation Details

Here we use analysis of synthetic data to supplement the discussion in Section 1 and further explore the properties of the proposed models. The dataset comprises 2000 locations distributed randomly within a 1000 × 1000 unit square domain (see Figure 3). The data were split to facilitate model comparison; half used to fit the candidate models ($\mathscr{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$) and the other half, referred to as the holdout set ($\mathscr{S}_0 = \{\mathbf{s}_{01}, \mathbf{s}_{02}, \ldots, \mathbf{s}_{0m}\}$), used to validate the models' predictive ability. The outcome variable associated with all locations was generated using the likelihood in (3). When an outcome variable shows spatial structure, it is common that both observed and unobserved predictor variables will as well. For example, this is seen in the subsequent NFI data analysis and illustrated in Figures 1 and 2. We therefore construct the synthetic data such that the outcome variables, $y(\mathbf{s})$ and $z(\mathbf{s})$, are partially described by the spatial predictors in $\mathbf{X}$ and $\mathbf{V}$, respectively. Specifically, $\mathbf{X}$ and $\mathbf{V}$ are $n \times 2$ matrices with an intercept as their first columns, while the second columns were drawn from $MVN\left(\boldsymbol{\mu}_x, \sigma_x^2 \mathscr{R}(\phi_x)\right)$ and $MVN\left(\boldsymbol{\mu}_\nu, \sigma_\nu^2 \mathscr{R}(\phi_\nu)\right)$, respectively. The (i, j)th element in the intersite correlation matrix R equals $\rho(\mathbf{s}_i, \mathbf{s}_j; \cdot)$, where the correlation function is given by (2). For simplicity, we fixed the smoothness parameter $\boldsymbol{\phi}_2$ for both predictor variables at 0.5 which reduces (2) to an Exponential correlation function. Other parameter values were set to $\boldsymbol{\mu}_x = 10$, $\sigma_x^2 = 1$, $\phi_x = 0.006$, $\boldsymbol{\mu}_\nu = 0$, $\sigma_\nu^2 = 10$, and $\phi_\nu = 0.006$.

Similarly, **w** and **u** were drawn from $MVN\left(\mathbf{0}, \sigma^2 \mathscr{R}(\phi)\right)$ and $MVN\left(\mathbf{0}, \sigma_{\mathbf{z}}^2 \mathscr{R}(\phi_{\mathbf{z}})\right)$, respectively. These and other synthetic data parameter values are given in Table 1. The subfigures in Figure 4 illustrate how the values of the spatial predictor (a) and the spatial effects (b), influence the probability (c) of a measurable outcome. This probability is imposed on the continuous process which is described by its spatial predictor (d) and spatial effects (e). Again, these processes are coupled through (3) to determine the value of the outcome variable across the domain, (f).

For this analysis we consider three candidate models. All use the predictive process to reduce the dimensionality of the spatial random effects and were fit using knot intensities of 49, 100, and 144. The first, a Gaussian spatial regression, is given by the first stage density in (3) with $z(\mathbf{s})$ fixed at 1 (i.e., a common geostatistical model used for continuous outcome variables). We refer to this as the *direct model* because it assumes all $y(\mathbf{s})$ come from the same Gaussian distribution (i.e., both zero and nonzero values). For fairness, the design matrix of this regression includes an intercept and the spatial predictor in **X** and **V**. The second and third are the full hierarchical form (3), with the distinction that one uses the nonbias-adjusted and the other uses the bias-adjusted predictive process for **w** and **u**. These candidate models were assessed based on their ability to recover the parameters used to generate the data and subsequent accuracy and precision of prediction at the holdout set locations.

In addition, we consider a second analysis using the synthetic data. Here we are interested in the models' ability to predict the holdout set total $y_{\mathscr{S}_0} = \sum_{i=1}^{m} y(\mathbf{s}_{0i})$. As noted in Section 1, the common approach to quantifying a forest variable over an area of interest is to discretize the predicted probability of forest occupancy at all locations, then predict the forest variable of interest at only locations deemed forested. These predictions, or samples from the posterior predictive distributions, can then be summed to obtain the total posterior predictive distribution of the forest variable over the set of locations that define the area. Correctly predicting $y_{\mathscr{S}_0}$ requires the $z(\mathbf{s}_{0i})$'s to be known. Otherwise, if $z(\mathbf{s}_{0i})$'s are not known, the associated uncertainty should be reflected in the width of the credible intervals for the $y_{\mathscr{S}_0}$. This can be illustrated using the direct and full hierarchical models, described previously, as well as the common *two-step* approach.

For the two-step approach, $z(\mathbf{s}_i)$'s and associated **V** were used to fit a logistic regression with spatial random effects. The parameter estimates from this model were then used to predict the probability of forest at each holdout location. These probabilities were then discretized at 0.5 to identify the predicted forest status of the holdout locations. A subsequent Gaussian spatial regression model was fit using only forested observations [i.e., $y(\mathbf{s}_i)$'s where $z(\mathbf{s}_i) = 1$] and their respective predictors in **X**. This model was then used to predict $y_{\mathscr{S}_0}$. Also, for comparison, we generated a fourth set of predictions using the two step approach but assumed the logistic regression predicted $z(\mathbf{s}_{0i})$'s without classification error.

To complete the Bayesian specification, we assign priors to the models' parameters. Customarily, we assume the $\boldsymbol{\beta}$'s and $\boldsymbol{\eta}$'s have *flat* prior distributions and the τ^2, σ^2, and σ_z^2 follow an inverse-Gamma (IG) with hyperparameters IG(2, 1). Note, with a shape value of

2, the IG distribution has infinite variance and is centered on the scale value, which in this case is 1. We experimented with a range of scale values but saw negligible change in the final parameter estimates. The spatial decay parameters ϕ and ϕ_z each follow a Uniform prior $U(0.003, 3)$, which corresponds to support for an effective spatial range (i.e., the distance at which the spatial correlation drops to 0.05) from 1 to 1000.

Three MCMC chains, with unique starting values, were run for 200,000 iterations. The CODA package in R (www.r-project.org) was used to diagnose convergence by monitoring mixing using Gelman–Rubin diagnostics and autocorrelations (see, e.g., Gelman et al. 2004, section 11.6). Acceptable convergence was diagnosed within 50,000 iterations (which were discarded as burn-in). The sampler was coded in C++ and Fortran and leveraged Intel's Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. The code was run on a Linux workstation with two Intel Nehalem-based quad-Xeon processors.

4.2 Results and Discussion

Post burn-in posterior summary statistics for each parameter are offered in Table 1. Bolded values indicate those 95% credible intervals that do not cover the *true* parameter values. For brevity, we only present estimates for the 49 and 144 knot intensities for the predictive-process models. Here we see the direct model yielded enormously inflated estimates of τ^2 and σ^2 which reflects its inability to separate the continuous and binary processes. As an aside, a more traditional approach would begin by fitting an empirical semivariogram to the residuals of the nonspatial Gaussian regression model, Figure 5. Then, one would use the estimates of τ^2 (i.e., *nugget*) and σ^2 (i.e., *partial sill*) to approximate the spatial dependence among the observations. As expected, the empirical semivariogram produces parameter estimates similar to the direct model. Moving to the hierarchical models' parameter estimates, the upward bias incurred by using the nonbias-adjusted predictive process is only seen in estimates of τ^2. Also, the non bias-adjusted predictive process model slightly underestimates η_1. These discrepancies, however, are corrected with the bias adjustment.

Each candidate model was used to predict the outcome value for the 1000 holdout locations. The posterior distribution of root mean squared error (RMSE) (i.e., the square root of the mean squared difference between the observed and predicted values) and bias (i.e., mean difference between observed and predicted values) were then estimated using the samples from each models' posterior predictive distribution. These values are given in Table 2. Here, we see the direct model produces nearly double the RMSE of the hierarchical models. Further, despite the discrepancy in parameter estimates of τ^2 and η_1, there was marginal difference in RMSE and bias produced by the two hierarchical models (e.g., posterior median and 95% credible interval of 13.37 (12.11, 14.56) and 12.10 (10.97, 13.12) for the 144 knot nonbias-adjusted and bias-adjusted predictive process models, respectively). None of the models produce a prediction bias significantly different from zero.

Figure 6 helps us understand the disparity in the predictive performance between these models and further corroborates the inadequacy of the direct model. Here, the solid line represents the aspatial density of the holdout set's outcome variable. The hierarchical model is able to use the spatial predictor variables and spatial random effects to appropriately apportion the predictions between the mass at zero and the measurable values of the

outcome. In contrast, the direct model produces an erroneous unimodal distribution of predictions. The subfigures in Figure 7 present interpolated surfaces of the median posterior predictive distribution over the holdout locations based on 100 knots. The hierarchical model (c) produces a surface similar to the observed (a), whereas the direct model (b) again grossly oversmooths.

Table 3 provides the predicted total of the holdout set locations. The *true* total of the holdout set $y_{\mathscr{S}_0} = 22.6 \times 10^3$. Again we see the direct model performs poorly for predictive inference. The 95% prediction interval for $y_{\mathscr{S}_0}$ from the hierarchical model and two-step model, with $z(\mathbf{s}_{0i})$'s assumed known, are centered on the true total. However, by acknowledging uncertainty in predicting $z(\mathbf{s}_{0i})$'s, the hierarchical model produces marginally wider credible intervals, for example, for the 144 knot models the 95% intervals are (21.6, 23.1) versus (22.5, 22.7). In reality the status of $z(\mathbf{s}_{0i})$'s is unknown a priori, therefore the more realistic version of the two step model first predicts $z(\mathbf{s}_{0i})$'s [labeled as "$z(\mathbf{s}_{0i})$'s unknown" in Table 3]. In this case, the spatial logistic regression achieved a classification accuracy of ~94% for the 49, 100, 144 knot intensities. Obviously, V and the spatial effects help explain a large portion of the variability in the probability that $z(\mathbf{s}) = 1$. However, despite this small classification error, the 95% prediction interval produced using the two-step model does not include the true holdout set total $y_{\mathscr{S}_0}$ due to misclassification of the $z(\mathbf{s}_{0i})$'s.

5. ANALYSIS OF FOREST INVENTORY DATA

5.1 Implementation Details

Given the data described in Section 2 and methods detailed in Section 3 we consider three candidate models for predicting forest biomass at a 30 × 30 m resolution over the study area. The candidate models follow (3) and include the nonspatial (i.e., $\mathbf{w} = \mathbf{0}$ and $\mathbf{u} = \mathbf{0}$), spatially varying intercepts, and spatially varying coefficients forms. Models were fit using three knot intensities of 100, 200, and 300. Knot locations were chosen using a k-means clustering on the FIA plot locations. Figure 8(a) illustrates the resulting 200 knot configuration.

As in the synthetic analysis, we use a *flat* prior on all $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ parameters. For the univariate spatial processes in the spatially varying intercept model, priors must be specified for, σ^2 and σ_z^2, and the Matérn correlation function's decay, ϕ_1, and smoothness, ϕ_2, parameters. We assume that σ^2 and σ_z^2 follow an IG(2, 1). Again, with a shape value of 2, the IG distribution has infinite variance and is centered on the scale value, which in this case is 1. Evaluating the Bessel function in (2) for each element in $\boldsymbol{\Sigma}_{\mathbf{w}}$ and $\boldsymbol{\Sigma}_{\mathbf{u}}$ in every MCMC iteration requires significant computing time. Therefore, we opt to fix ϕ_2 and $\phi_{z,2}$ at 0.5, which reduces the Matérn to the familiar exponential correlation function. Given the maximum intersite distance of 489.08 km, we assigned ϕ and ϕ_z a uniform prior with broad spatial support $U(5 \times 10^{-6}, 1)$, which is between 1 and 600,000 m. For the spatially varying coefficients model, we assumed the $p \times p$ and $q \times q$ dimensional cross-covariance matrices $\mathbf{L}\mathbf{L}'$ and $\mathbf{L}_z\mathbf{L}_z'$ follow a IW($p + 1$, I_p) and IW($q + 1$, I_q) prior, respectively. We experimented with different diagonal IW's scale matrices to assess this prior's influence on posterior distributions an found that as long as these values were small, there was negligible change in

the parameter estimates. The spatial decay parameter associated with each β and η specific process again follow $U(5 \times 10^{-6}, 1)$.

The models were fit in the computing environment described in Section 4. For each model, three chains were run for 200,000 MCMC iterations. Acceptable convergence was diagnosed within 50,000 iterations and therefore 450,000 samples (3 × 150,000) were retained for posterior analysis. The spatially varying coefficients model was the most computationally challenging with each chain of the 300 knot model taking ~72 hours to complete.

Because our primary interest is prediction, we compare the candidate models' ability to predict biomass for a set of $m = 877$ (i.e., 10%) holdout (or validation) plots that were selected at random from the 8774 FIA plots. Then, following (7), posterior predictive samples from each holdout set location were used to calculate the posterior distribution of RMSE and a deviance scoring rule. Here deviance of the holdout set using the kth posterior sample is

$$\begin{aligned} -2 \sum_{i=1}^{m} & -\frac{1}{2} log\left(z\left(\mathbf{s}_{0i}\right)\boldsymbol{\tau}_1^{2(k)} + \left(1 - z\left(\mathbf{s}_{0i}\right)\right)\boldsymbol{\tau}_2^{2(k)}\right) \\ & - \frac{\left(y(\mathbf{s}_{0i}) - \mathbf{x}(\mathbf{s}_{0i})'\boldsymbol{\beta}^{(k)} - \tilde{\mathbf{x}}(\mathbf{s}_{0i})'\mathbf{w}(\mathbf{s}_{0i})^{(k)}\right)^2 z(\mathbf{s}_{0i})}{2\left(z(\mathbf{s}_{0i})\boldsymbol{\tau}_1^{2(k)} + (1 - z(\mathbf{s}_{0i}))\boldsymbol{\tau}_2^{2(k)}\right)} \\ & + z\left(\mathbf{s}_{0i}\right) log\left(\frac{1}{1 + exp\left(-\mathbf{v}(\mathbf{s}_{0i})'\boldsymbol{\eta}^{(k)} - \tilde{\mathbf{v}}(\mathbf{s}_{0i})'\mathbf{u}(\mathbf{s}_{0i})^{(k)}\right)}\right) \\ & + \left(1 - z\left(\mathbf{s}_{0i}\right)\right) \times log\left(1 - \frac{1}{1 + exp\left(-\mathbf{v}(\mathbf{s}_{0i})'\boldsymbol{\eta}^{(k)} - \tilde{\mathbf{v}}(\mathbf{s}_{0i})'\mathbf{u}(\mathbf{s}_{0i})^{(k)}\right)}\right), \end{aligned}$$

where $\mathbf{s}_{0i}$'s denote the hold-out locations.

In addition to the holdout set analysis, we also compare the proposed model results of total biomass over multipixel areas to those obtained using design-based inference. As noted in the introduction, design-based inference that uses asymptotically unbiased estimators is the standard, albeit limited, approach to summarizing forest variables. Given a simple random sample of n_A plots within an area of interest the design-based estimator for the mean and variance of predicted biomass is $\hat{y}_A = \sum_{i=1}^{n_A} y_i / n_A$ and $var\left(\hat{y}_A\right) = \sum_{i=1}^{n_A}\left(y_i - \hat{y}_A\right)^2 / n_A$. Using at least 30 plots, these estimators should provide reasonable representation with associated uncertainty of biomass. Therefore, we selected at random small areas of interest containing 30 FIA plots [Figure 9(a)]. The design-based and proposed model-based predictions of total biomass for each area of interest were then computed.

5.2 Results and Discussion

Table 4 offers parameter estimates for the three candidate models. All models at the 300 knot intensity have statistically significant regression coefficients associated with Normalized Difference Vegetation Index (NDVI), precipitation (PRECIP), and minimum average annual temperature (TMIN). Also, the three tasseled cap indices explain a significant portion of the variation in the probability of forest. Although it is not our central

interest here, we are often interested in a physical interpretation of the predictor variables' coefficients in relation to the outcome variable. If this is the focus, then significance of a predictor variable's coefficient in the spatially varying coefficient models must be done in a spatial context. That is, rather than look at the aspatial β and η, which for these models are simply the mean over the domain, interpretation should be based on surfaces of **w** and **u**.

Although small relative to the residual variance τ^2, the spatial processes in the spatially varying intercept and coefficients models capture a substantial portion of the variability in forest biomass and probability of forest (i.e., the diagonal elements of $\mathbf{LL}'$ and $\mathbf{L}_z\mathbf{L}_z'$, respectively). The values of the spatial decay parameters ϕ and ϕ_z also support the use of a spatial model. For instance, the median estimated effective spatial ranges and associated 95% credible intervals in km for the 300 knot spatially varying coefficients model's ϕ and ϕ_z are 352 (290, 429), 279 (242, 322), 178 (139, 234), 326 (298, 368), and 292 (261, 319) for the intercept, NDVI PRECIP, TMAX, and TMIN, respectively, and 369 (340, 432), 264 (201, 294), 357 (296, 403), 399 (365, 465) for the intercept, TC1, TC2, and TC3, respectively. These long spatial ranges help support our initial simplifying assumption to combine the data layers into a common pixel resolution.

We now turn to the results of the 877 plot holdout set analysis summarized in Table 5. For both the deviance and RMSE scoring rules a lower value indicates improved predictive performance. Not surprisingly, accounting for missing predictor variables by adding a spatial random effect to the model's intercepts improves prediction, that is, the median deviance of the nonspatial model is −3,796.52 and the 100 knot spatially varying intercept is −3,818.42. Further, allowing the regression coefficients to vary spatially also improves prediction. This is shown by a statistically significant separation of the posterior distributions of the deviance score between the two spatial models across knot intensities. Table 5 also shows a progressively smaller RMSE with increasing model richness, that is, medians of 7.12, 6.81, and 6.43 for the nonspatial and 300 knot spatially varying intercept and coefficients models, respectively. Together, these results support the use of a spatially varying coefficients model. Interpolated surfaces of this model's fitted values are illustrated in Figure 8(b) and (c). As expected, these surfaces approximate the interpolated surfaces of the observed data in Figure 1. It is likely that increasing knot intensity beyond 300 knots will further improve the predictive performance; however, the run time under the current computing environment becomes unreasonable.

Using the proposed hierarchical models, we can build data layers that allow end users to assess the outcome variable of interest, associated uncertainties, and to propagate error through subsequent models at a fine spatial resolution. Using the 300 knot spatially varying coefficients model we made pixel-level predictions for each of the 50 small areas of interest illustrated in Figure 9(a). Each area consists of approximately 110,000 30 × 30 m pixels. Figure 10 illustrates two such areas with heterogeneous landscapes consisting of forest, water, agricultural fields, and developed land use (e.g., roads, power lines, urban). These features are apparent in Subfigures (a) and (b) which depict the mean of each pixel's posterior predictive distribution of forest biomass. Here, the scale bars indicate the pixel-level estimate of biomass in metric tons per ha. Further, Subfigures (b) and (d) show the

associated standard deviation of biomass prediction. As noted previously, unlike model-based approaches, the design-based inferences is typically limited to point estimates of a few first and second moments over areal units. Figure 9(b) offers the design-based mean and associated 95% confidence interval of total biomass for each of the 50 areas versus the mean of the posterior predictive distribution (note, the point closest to the origin does not have confidence bars because all FIA plots were nonforest). The one-to-one line is also plotted and passes through 48 of the 50 confidence interval which indicates the proposed model does not produce a systematic bias in predicted forest biomass within these areas; however, this argument is a bit tenuous. Rather, we use this plot to illustrate the reasonable level of correspondence between the two approaches.

The propensity to misclassify forest occupancy of a pixel, and utility of the proposed model, should increase with increasing landscape heterogeneity and pixel size. Globally, forests continue to become more fragmented due to agricultural and urban development; it has been estimated that a major portion of the world's tree cover is no longer in lands classed as forest (Ellis and Ramankutty 2008). Misclassification of pixels as forest or nonforest has been demonstrated to be high under such conditions (Sivanpillai et al. 2005; Fang et al. 2006). For example, in a recent study in East Texas, USA, forest classifications from moderate spatial resolution Landsat satellite imagery (30×30 m pixel) were found to range from about 81 to 91% accuracy in predicting observed forest areas across six counties, with a net overestimation of forested area due largely to misclassification of other treed spaces (e.g., urban areas and pastures with significant tree cover) (Sivanpillai et al. 2005). Misclassification error in forest area estimates tend to increase as pixel size increases, holding area constant, or as area decreases, holding pixel size constant (Zheng, Heath, and Ducey 2008a). While regional and global scale observations of forest variables are of great interest they often require that larger pixel sizes be used and hence the opportunity for greater uncertainty in the forest variable of interest due to forest/nonforest misclassification error (Zheng, Heath, and Duce 2008b).

6. SUMMARY

Our central interest was to propose and explore a hierarchical modeling framework for predicting forest variables over large heterogeneous domains. Our framework addresses uncertainty in both forest occupancy and the forest variable of interest by incorporating predictor variables and spatial random effects at each level of the hierarchy. Specification of a multivariate spatial process affords additional flexibility by allowing each predictor variable's regression coefficient to vary spatially across the domain, which, in turn, can improve predictive performance. To handle the realistically large number of locations in a hierarchical setup we adopt a recently proposed reduced-rank process referred to as the predictive process. Importantly, the hierarchical structure provides a mechanism to propagate error which ultimately provides a more realistic expression of uncertainty in the predicted outcome. It would be difficult to estimate the spatial parameters a priori, and even if reasonable estimates could be made, the *plug-in* approach used in traditional methods can provide falsely precise estimates of predicted outcome. This could, in turn, negatively impact end users' sensitivity analyses.

It is worth pointing out that one could envisage alternative modeling approaches based upon a nonlinear function that assigns low or zero expected biomass when covariates suggest little or no forest cover. This could, potentially, lead to a more “direct” one-stage regression model that may be attractive by virtue of being simpler and, perhaps, computationally faster. Such an approach, however, will not be without problems. Fore-most among them is the fact that predictors for biomass, and similar forest variables, can be very different from those that accurately predict forest occupancy. This may considerably compound the problem of arriving at a single nonlinear function that will behave appropriately and offer sound predictive performance. Furthermore, accommodating spatially varying coefficients will considerably increase these difficulties. These naturally lead to further questions about choosing between parametric and non-parametric options for modeling such nonlinear functions. The former could easily efface some of the spatial structure we want the spatial process to capture, thereby leading to poorer inference, while the latter would raise questions about choices of basis functions and could become especially unwieldy when the impact of the predictors is posited to vary across space. Nevertheless, these approaches, or modifications thereof, compose projects worthy of future investigations.

There is global interest in quantifying forest carbon sources and sinks from an economic and environmental perspective. Developing models to predict carbon proxies such as forest biomass using dense inventory plot arrays and spatially complete sets of environmental predictor variables is a useful exercise. Our results suggest the hierarchical model with spatially varying coefficients is well suited to this objective. Further, a sampling-based Bayesian approach to model fitting offers estimates of the spatial parameters and access to posterior predictive distributions for each new location. The biomass analysis presented here uses only a few coarse spatial resolution predictor variables. Undoubtedly, incorporating fine-scale forest and tree level variables through functions that model physical ecological processes will greatly improve predictions. We hope to extend this model by developing these suites of variables and complimentary dynamic functions to model mean response at each level in the hierarchy.

Finally, analysis of large spatial domains is becoming more common due, in part, to increased access to high performance mathematical libraries and computers and improvements in dimension reduction methods such as low-rank spatial processes. Only through a combination of these tools was the analysis presented here computationally feasible. With expanding domains of interest comes an increasing propensity for nonstationarity in the underlying spatial process. From a statistical validity standpoint, it is important that we define models that are equipped to deal with nonstationarity. From a utilitarian perspective, and as seen in our results, addressing nonstationarity can improve model fit and, more importantly, prediction.

Acknowledgments

This work was supported by Michigan AgBioResearch, the USDA Forest Service Forest Inventory and Analysis National Program, Forest Health Technology Enterprise Team, National Science Foundation DMS-0706870, and the Statistics and Applied Mathematical Science Institute.

REFERENCES

Albert, DA. Regional Landscape Ecosystems of Michigan, Minnesota, and Wisconsin: A Working Map and Classification. USDA Forest Service, North Central Forest Experiment Station; St. Paul, MN: 1995. General Technical Report NC-17832

Banerjee S, Gelfand AE, Finley AO, Sang H. Gaussian Predictive Process Models for Large Spatial Datasets. Journal of the Royal Statistical Society, Ser. B. 2008; 70:825–848. 36.

Bechtold, WA.; Patterson, PL. The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures. USDA Forest Service, Southern Research Station; Asheville, NC: 2005. p. 85General Technical Report SRS-8032

Crainiceanu CM, Diggle PJ, Rowlingson B. Bivariate Binomial Spatial Modeling of Loa Loa Prevalence in Tropical Africa. Journal of the American Statistical Association. 2008; 103:21–37. with discussion. 36,37.

Cressie, N. Statistics for Spatial Data. 2nd ed.. Wiley; New York: 1993. 32, 34

Cressie N, Johannesson G. Fixed Rank Kriging for Very Large Spatial Data Sets. Journal of the Royal Statistical Society, Ser. B. 2008; 70:209–226. 36.

Daly C, Taylor GH, Gibson WP, Parzybok TW, Johnson GL, Pasteris PA. High-Quality Spatial Climate Data Sets for the United States and Beyond. Transactions of the American Society of Agricultural and Biological Engineers. 2000; 43:1957–1962. 32.

Daniels MJ, Kass RE. Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. Journal of the American Statistical Association. 1999; 94:1254–1263. 35.

Diggle PJ, Lophaven S. Bayesian Geostatistical Design. Scandinavian Journal of Statistics. 2006; 33:53–64. 37.

Diggle PJ, Tawn JA, Moyeed RA. Model-Based Geostatistics. Applied Statistics. 1998; 47:299–350. with discussion. 31.

Dong J, Kaufmann RK, Myneni RB, Tucker CJ, Kauppi PE, Liski J, Buermann W, Alexeyev V, Hughes MK. Remote Sensing Estimates of Boreal and Temperate Forest Woody Biomass: Carbon Pools, Sources, and Sinks. Remote Sensing of Environment. 2003; 84:393–410. 32.

Ellis EC, Ramankutty N. Putting People in the Map: Anthropogenic Biomes of the World. Frontiers in Ecology and the Environment. 2008; 6:439–447. 46.

Fang S, Gertner G, Wang G, Anderson A. The Implication of Misclassification in Land Use Maps in the Prediction of Landscape Dynamics. Landscape Ecology. 2006; 21:233–242. 46.

Finley AO, Banerjee S, Ek AR, McRoberts RE. Bayesian Multivariate Process Modeling for Prediction of Forest Attributes. Journal of Agricultural, Biological, and Environmental Statistics. 2008b; 13:60–83. 32.

Finley AO, Banerjee S, McRoberts RE. A Bayesian Approach to Quantifying Uncertainty in Multi-Source Forest Area Estimates. Environmental and Ecological Statistics. 2008a; 15:241–258. 31.

Finley AO, Banerjee S, McRoberts RE. Hierarchical Spatial Models for Predicting Tree Species Assemblages Across Large Domains. The Annals of Applied Statistics. 2009a; 3:1052–1079. 32. [PubMed: 20352037]

Finley AO, Sang H, Banerjee S, Gelfand AE. Improving the Performance of Predictive Process Modeling for Large Datasets. Computational Statistics and Data Analysis. 2009b; 53:2873–2884. 37. [PubMed: 20016667]

Fuentes M. A New Class of Nonstationary Spatial Models. Biometrika. 2002; 89:197–210. 35.

Gaspari G, Cohn SE. Construction of Correlation Functions in Two and Three Dimensions. The Quarterly Journal of the Royal Meteorological Society. 1999; 125:723–757. 35.

Gelfand AE, Kim H, Sirmans CF, Banerjee S. Spatial Modeling With Spatially Varying Coefficient Processes. Journal of the American Statistical Association. 2003; 98:387–396. 32.

Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF. Nonstationary Multivariate Process Modeling Through Spatially Varying Coregionalization. Test. 2004; 13:263–312. with discussion. 35.

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. 2nd ed.. Chapman & Hall/CRC Press; Boca Raton, FL: 2004. 37,40

Gneiting, T.; Kleiber, W.; Schlather, M. Matérn Cross-Covariance Functions for Multivariate Random Fields. University of Washington, Dept. of Statistics; 2009. Technical Report 54935

Grzebyk, M.; Wackernagel, H. Multivariate Analysis and Spatial/Temporal Scales: Real and Complex Models; Proceedings of the XVIIth International Biometrics Conference; Hamilton, Ontario. 1994; p. 19-33.35

Harville, DA. Matrix Algebra From a Statistician's Perspective. Springer; New York: 1997. 35

Heagerty PJ, Lele SR. A Composite Likelihood Approach to Binary Spatial Data. Journal of the American Statistical Association. 1998; 93:1099–1111. 31.

Henderson HV, Searle SR. On Deriving the Inverse of a Sum of Matrices. SIAM Review. 1981; 23:53–60. 37.

Higdon, DM. Space and Space–Time Modeling Using Process Convolutions. In: Anderson, C.; Barnett, V.; Chatwin, PC.; El-Shaarawi, AH., editors. Quantitative methods for current environmental issues. Springer-Verlag; London: 2002. p. 37-56.36

Higdon DM, Gattiker J, Williams B, Rightley M. Computer Model Calibration Using High-Dimensional Output. Journal of the American Statistical Association. 2008; 103:570–583. 35.

Higdon, DM.; Lee, H.; Holloman, C. Markov Chain Monte Carlo-Based Approaches for Inference in Computationally Intensive Inverse Problems; Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting; Oxford: Oxford University Press. 2003; p. 181-197.with discussion35

Higdon, DM.; Swall, J.; Kern, J. Non-Stationary Spatial Modeling; Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting; Oxford: Oxford University Press. 1999; p. 761-768.35

Huang C, Wylie B, Homer C, Yang L, Zylstra G. Derivation of a Tasseled Cap Transformation Based on Landsat 7 At-Satellite Reflectance. International Journal of Remote Sensing. 2002; 8:1741–1748. 33.

Kaufman, L.; Rousseeuw, PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley; New York: 1990. 37

Lin X, Wahba G, Xiang D, Gao F, Klein R, Klein B. Smoothing Spline ANOVA Models for Large Data Sets With Bernoulli Observations and the Randomized GACV. The Annals of Statistics. 2000; 28:1570–1600. 37.

Majumdar A, Gelfand AE. Multivariate Spatial Modeling for Geostatistical Data Using Convolved Covariance Functions. Mathematical Geology. 2007; 39:225–245. 35.

Opsomer DJ, Breidt J, Moisen GG, Kauermann G. Model-Assisted Estimation of Forest Resources With Generalized Additive Models. Journal of the American Statistical Association. 2007; 102:400–416. 31.

Rasmussen, CE.; Williams, CKI. Gaussian Processes for Machine Learning. The MIT Press; Cambridge, MA: 2006. 36

Reich BJ, Fuentes M. A Multivariate Nonparametric Bayesian Spatial Framework for Hurricane Surface Wind Fields. The Annals of Applied Statistics. 2007; 1:249–264. 35.

Robert, CP.; Casella, G. Monte Carlo Statistical Methods. 2nd ed.. Springer; New York: 2005. 32

Royle JA, Nychka D. An Algorithm for the Construction of Spatial Coverage Designs With Implementation in SPLUS. Computers and Geosciences. 1998; 24:479–488. 37.

Ruppert, D.; Wand, MP.; Caroll, RJ. Semiparametric Regression. Cambridge University Press; Cambridge, U.K.: 2003. 37

Schmidt A, Gelfand AE. A Bayesian Coregionalization Model for Multivariate Pollutant Data. Journal of Geophysics Research-Atmospheres. 2003; 108:8783. 35.

Sivanpillai R, Smith CT, Srinivasan R, Messina MG, Wu XB. Estimating Regional Forest Cover in East Texas Using Enhanced Thematic Mapper (ETM+) Data. Forest Ecology and Management. 2005; 218:342–352. 46.

Stein, ML. Interpolation of Spatial Data: Some Theory of Kriging. Springer; New York: 1999. 32,35

Stein ML. Spatial Variation of Total Column Ozone on a Global Scale. The Annals of Applied Statistics. 2007; 1:191–210. 36.

Stein ML. A Modeling Approach for Large Spatial Datasets. Journal of the Korean Statistical Society. 2008; 37:3–10. 36.

Ver Hoef JM, Barry RD. Modelling Crossvariograms for Cok-riging and Multivariable Spatial Prediction. Journal of Statistical Planning and Inference. 1998; 69:275–294. 35.

Wackernagel, H. Multivariate Geostatistics: An Introduction With Applications. 3rd ed.. Springer-Verlag; New York: 2006. 35

Wahba, G. Spline Models for Observational Data. SIAM; Philadelphia: 1990. 36

Zheng D, Heath LS, Ducey MJ. Modeling Grain-Size Dependent Bias in Estimating Forest Area: A Regional Application. Landscape Ecology. 2008a; 23:1119–1132. 46.

Zheng D, Heath LS, Ducey MJ. Identifying Grain-Size Dependent Errors on Global Forest Area Estiamtes and Carbon Studies. Geophysical research letters. 2008b; 35:1–5. 46.

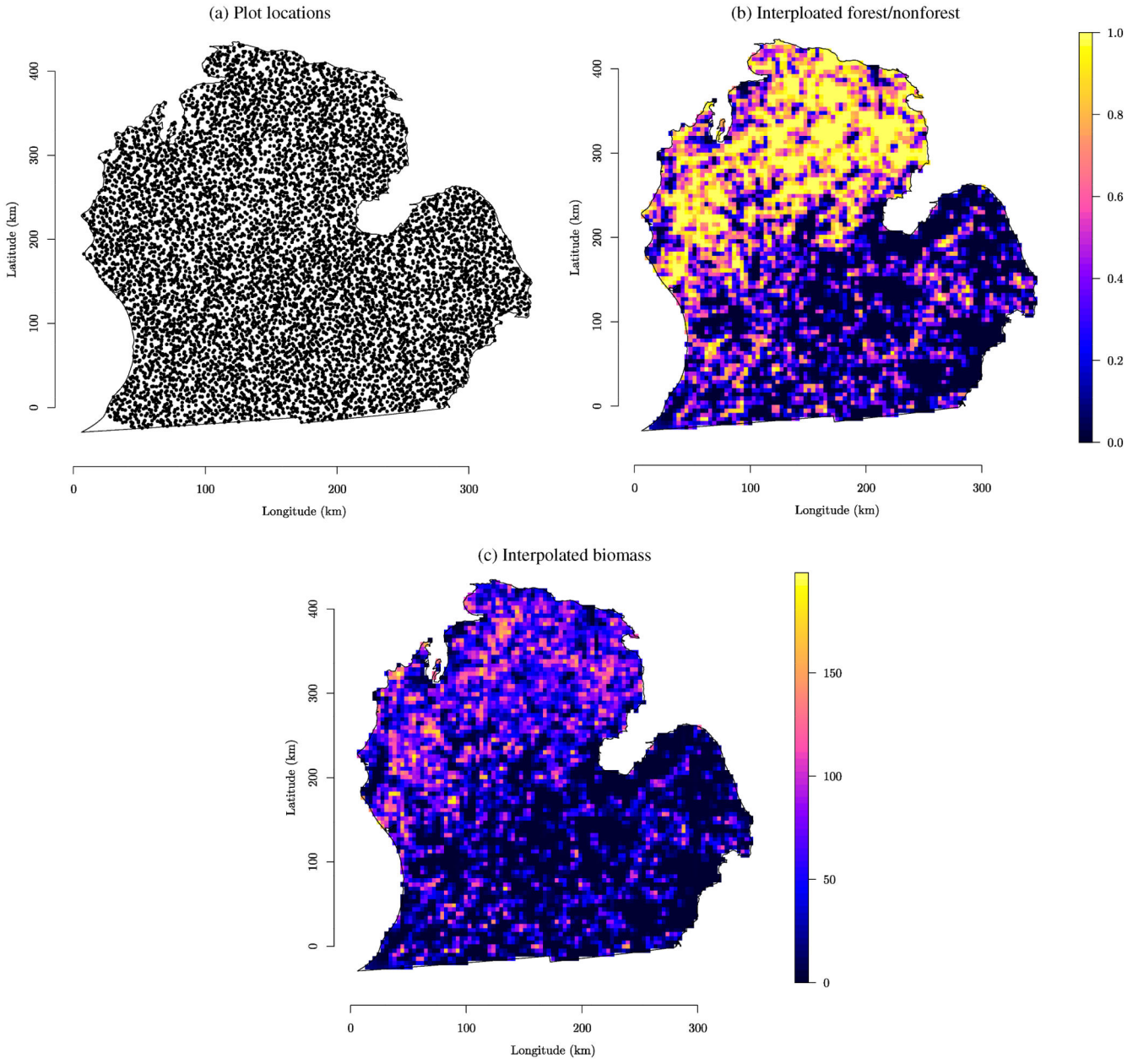

Figure 1.

Forest inventory plot locations (a), interpolated surface of biomass in metric tons per ha (c), and interpolated surface of forest/nonforest (b). In accordance with our Forest Service FIA data sharing agreement the plot locations in (a) are “fuzzed” to obscure the true locations. The online version of this figure is in color.

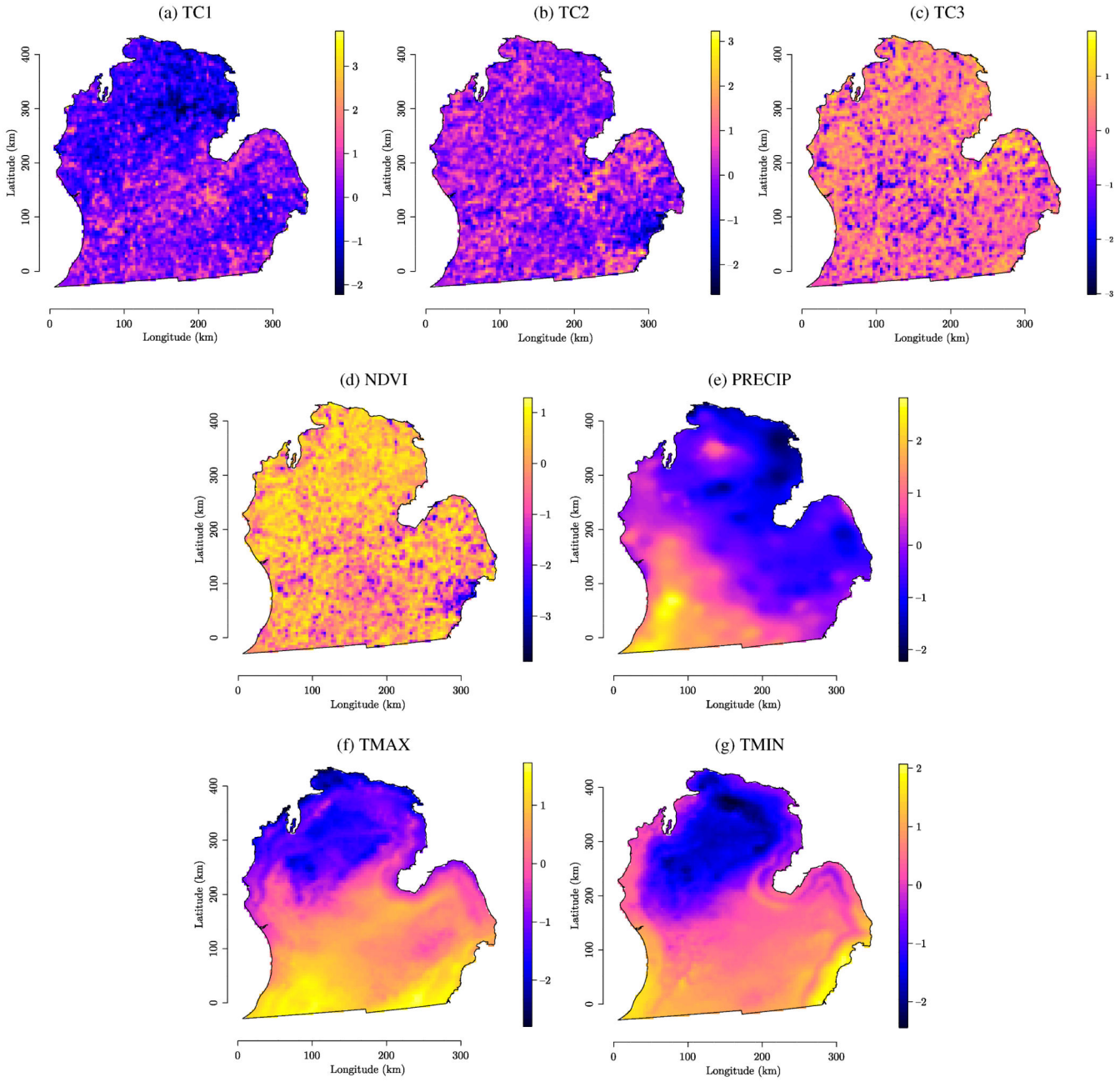

Figure 2.

Scaled predictor variables with (a)–(c) used for the binary process and (d)–(g) used for the continuous process. The online version of this figure is in color.

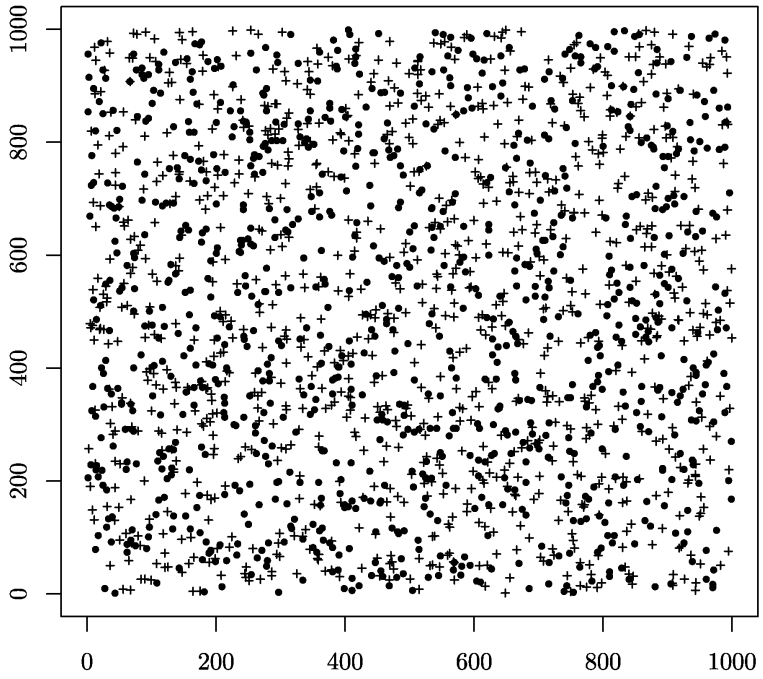

Figure 3.
Synthetic data locations used for model parameter estimation (•) and subsequent prediction (+).

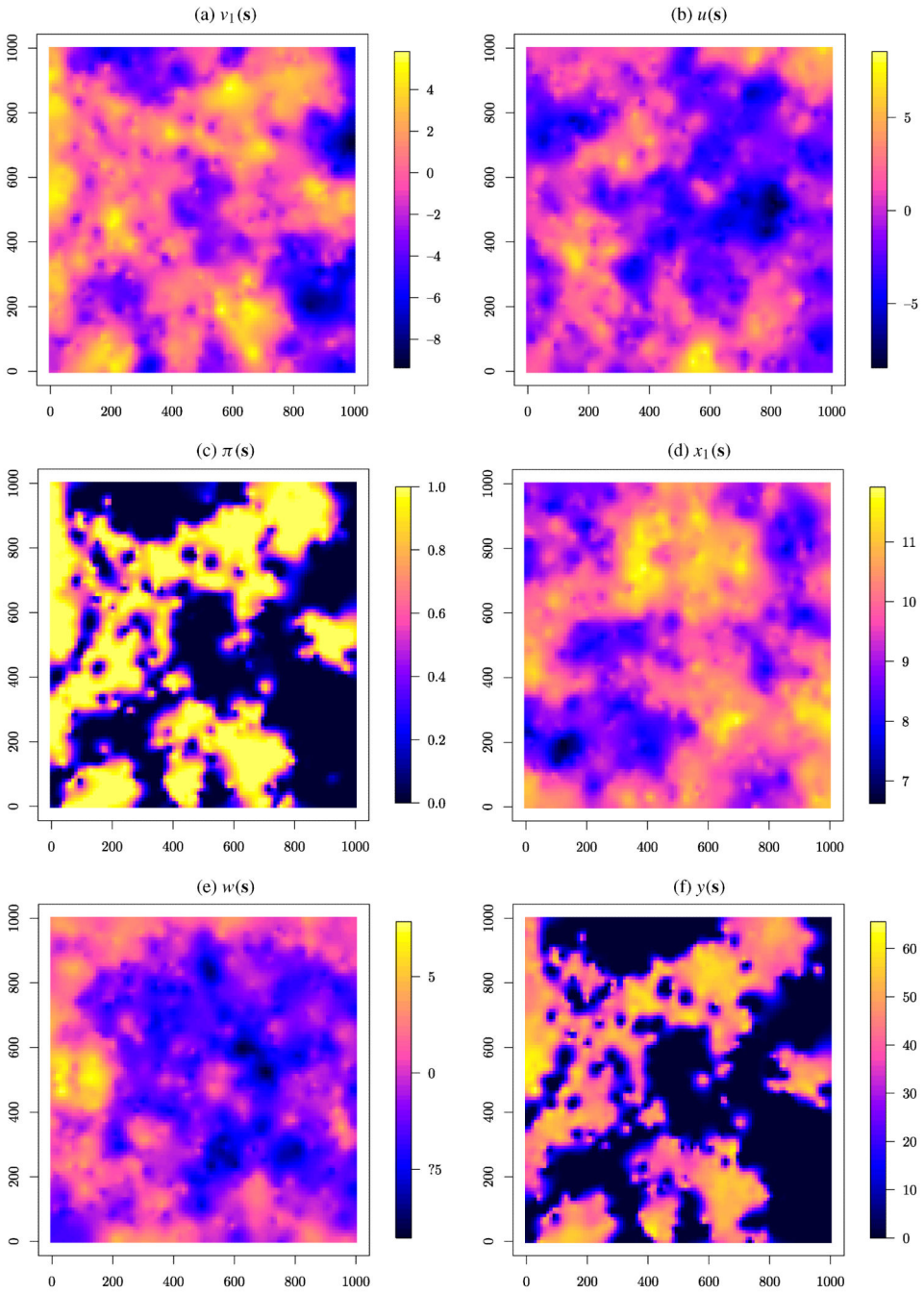

Figure 4.
Interpolated surfaces of the synthetic data quantities. The online version of this figure is in color.

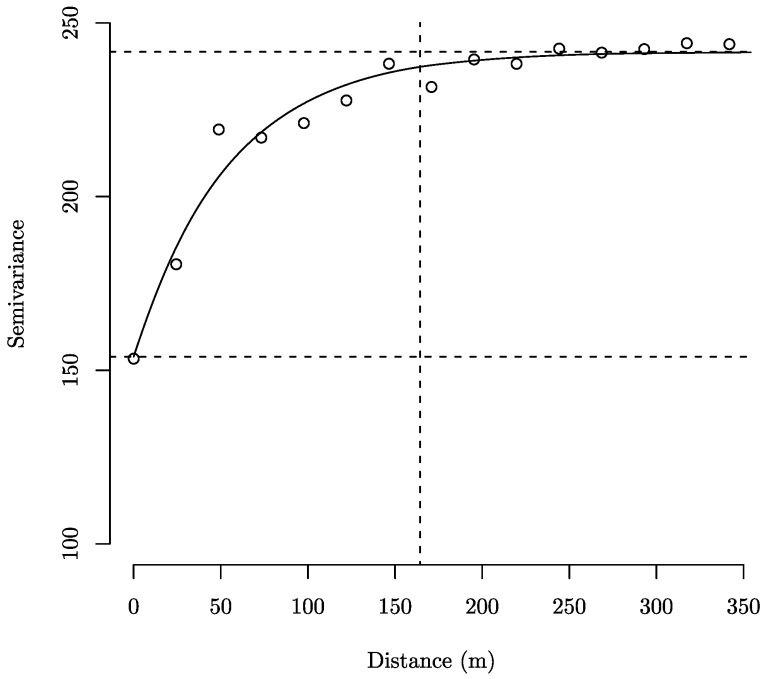

Figure 5.

Empirical semivariogram of the residuals from the synthetic data nonspatial candidate model. Maximum likelihood estimates of this exponential semivariogram's nugget, partial sill, and effective spatial range [i.e., $-\log(0.05/\phi)$ are indicated by the lower horizontal, upper horizontal, and vertical dotted lines, respectively].

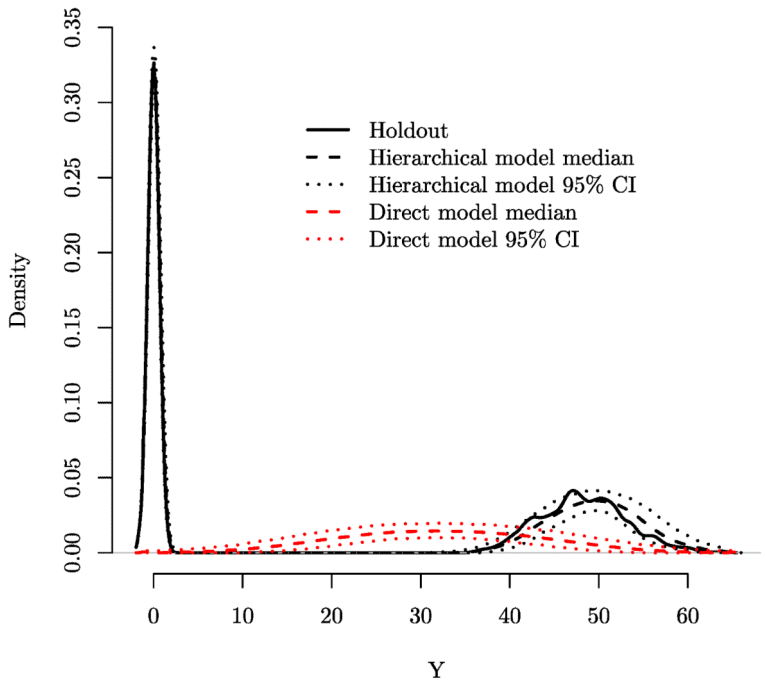

Figure 6.

Non-spatial distribution of the observed holdout set outcome variable $y(\mathbf{s}_{0i})$'s and associated predictions from the direct and hierarchical candidate models. The online version of this figure is in color.

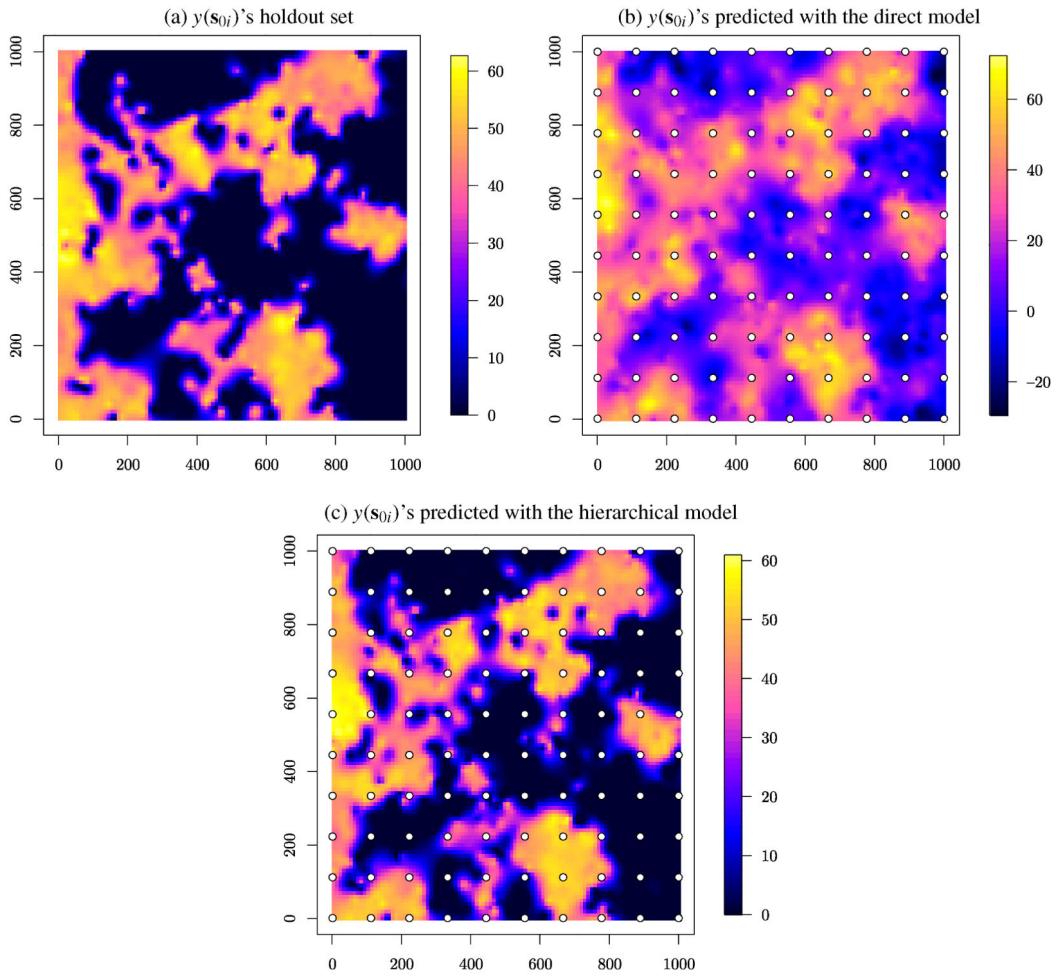

Figure 7.

Interpolated surface of the holdout set outcome variable $y(\mathbf{s}_{0i})$'s (a) and median of the posterior predictive distribution generated using the direct and hierarchical candidate models (b) and (c), respectively. The locations of the 100 knots are superimposed on (b) and (c). The online version of this figure is in color.

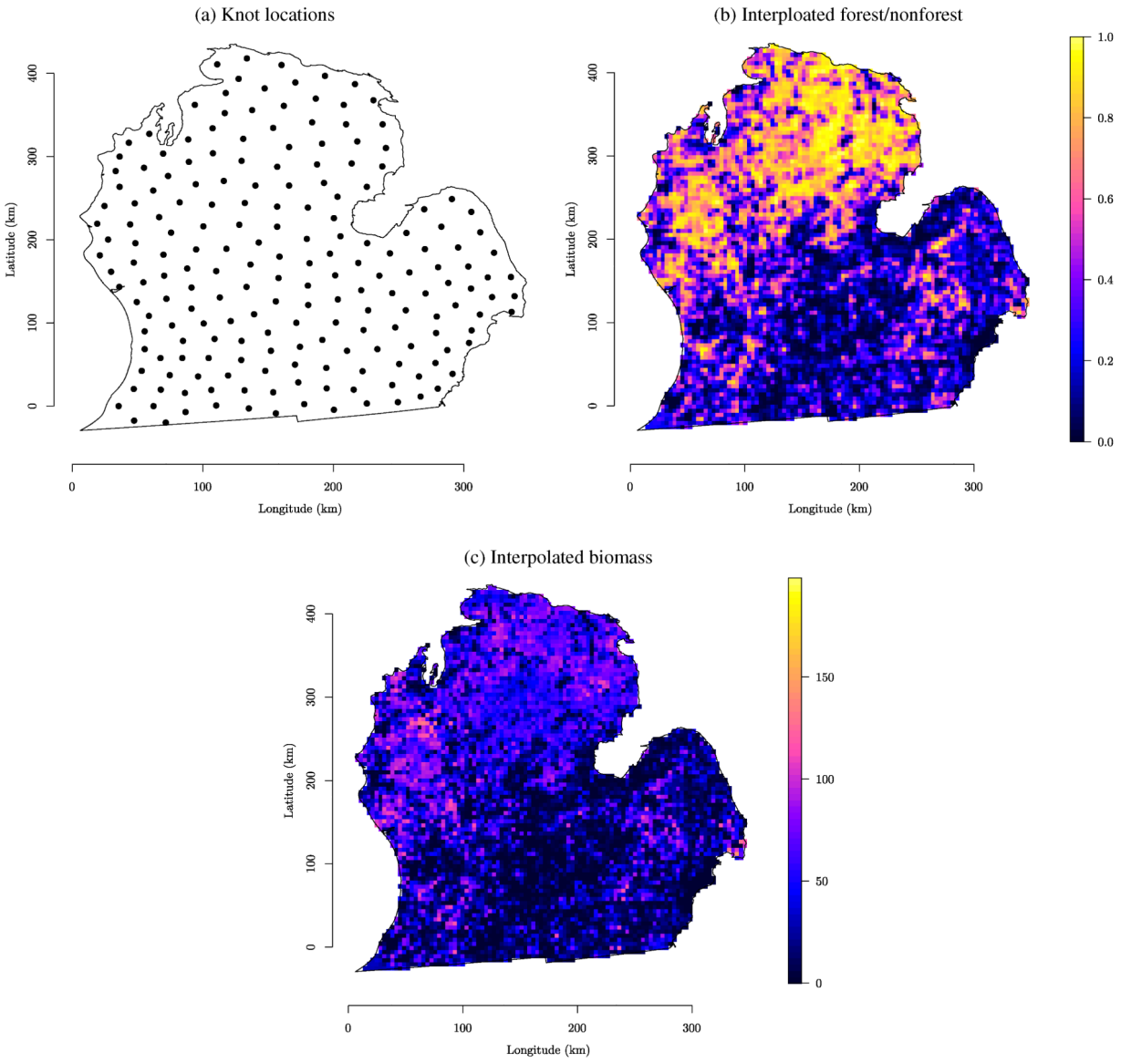

Figure 8.

Locations of 200 predictive process knots (a), interpolated surface of $\pi(\mathbf{s})$ forest/nonforest fitted values (b), and interpolated surface of biomass in metric tons per ha fitted values (c). The online version of this figure is in color.

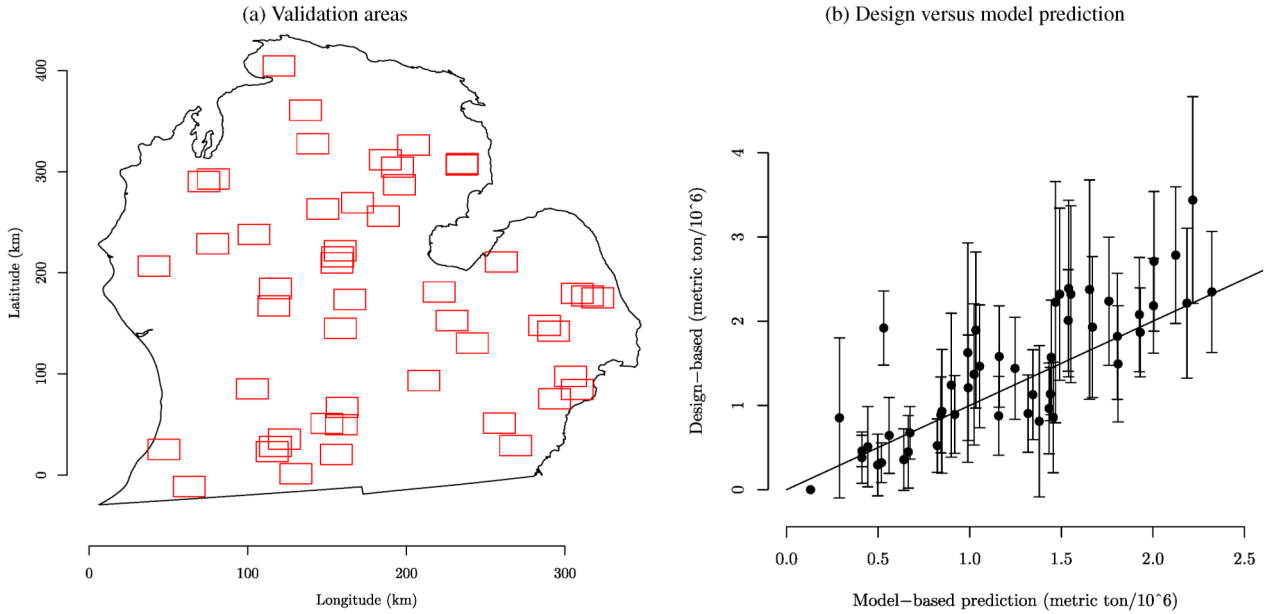

Figure 9.

Randomly selected areas used for validation (a) and design versus model-based estimates of total biomass for each validation area (b). Here the point and intervals are the design-based mean and 95% confidence intervals of total biomass (metric tons per ha) for each validation area. The one-to-one line is given for comparison of two methods. The online version of this figure is in color.

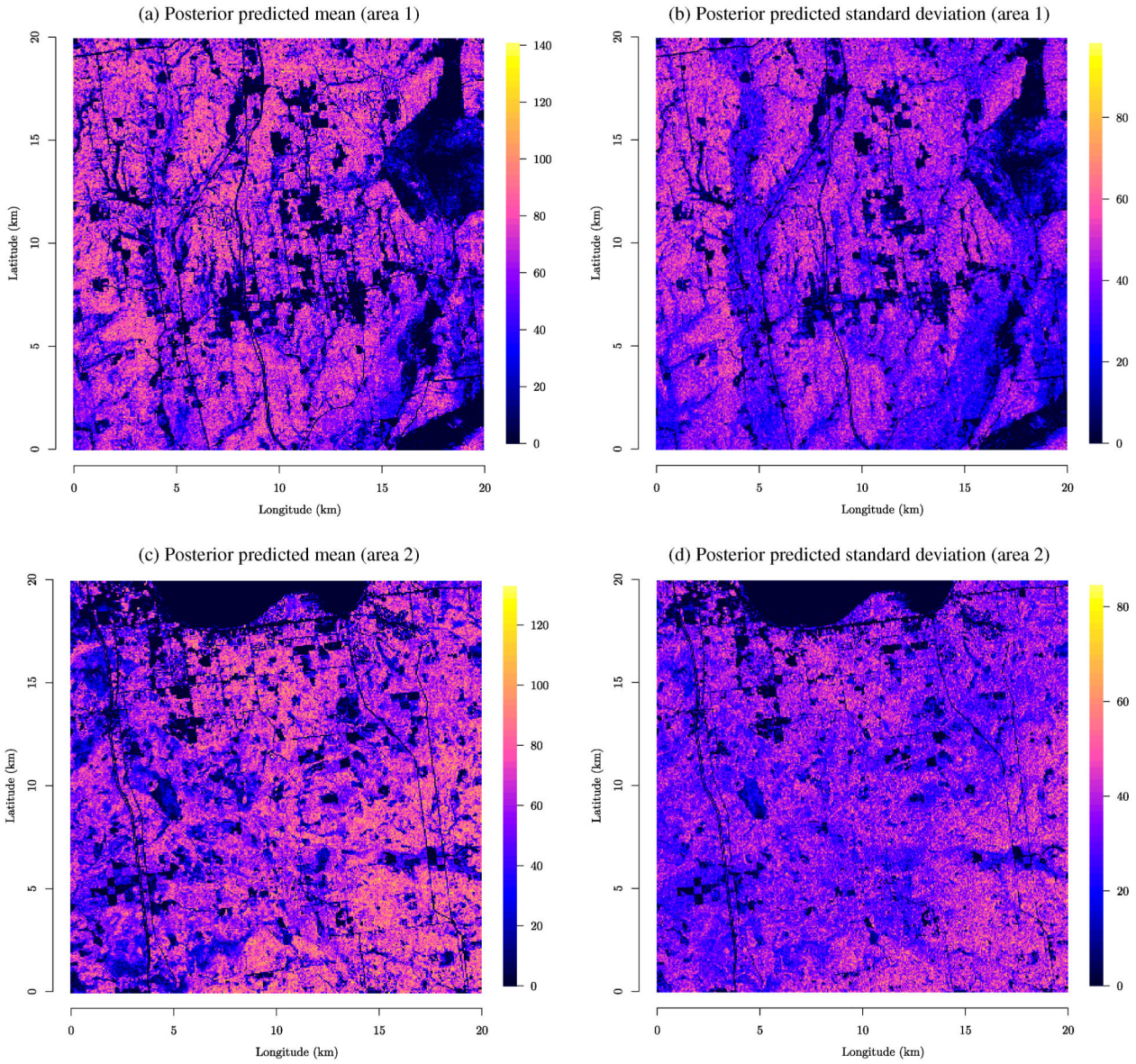

Figure 10.

Examples of validation areas' pixel-level posterior predictive distribution mean and standard deviation of biomass (metric tons per ha). The online version of this figure is in color.

Table 1

Parameter credible intervals, 50 (2.5, 97.5) percentiles, for synthetic data analysis candidate model

		Direct	Hierarchical model			
		Bias-adjusted	Nonbias-adjusted predictive process		Bias-adjusted predictive process	
Parameter	**True**	**144**	**49**	**144**	**49**	**144**
β_0	1	−8.28 (−23.87, 6.89)	3.35 (−0.23, 6.94)	2.04 (−1.55, 5.45)	−0.36 (−1.59, 1.21)	−0.26 (−1.57, 1.15)
β_1	5	3.73 (2.19, 5.31)	4.66 (4.34, 5.02)	4.75 (4.43, 5.09)	5.13 (4.96, 5.26)	5.13 (4.99, 5.26)
η_0	1	–	0.65 (−0.14, 1.58)	0.76 (−0.27, 2)	1.11 (0.74, 1.49)	1.16 (0.78, 1.54)
η_1	5	**7.29 (6.8, 7.8)**	**3.39 (2.87, 4)**	**3.88 (3.19, 4.68)**	5.26 (4.66, 5.97)	5.3 (4.69, 5.99)
τ^2	1	**187.05 (170.25, 206.03)**	**3.17 (2.78, 3.63)**	**2.54 (2.19, 2.96)**	1.21 (0.95, 1.54)	1.04 (0.84, 1.24)
σ^2	0	**120.33 (64.72, 201.37)**	10.69 (6.46, 18.57)	9.99 (6.29, 17.03)	8.05 (5.43, 14.09)	8.38 (5.82, 14.65)
ϕ	0.006	**0.015 (0.008, 0.023)**	0.007 (0.003, 0.012)	0.007 (0.004, 0.011)	0.005 (0.003, 0.007)	0.005 (0.003, 0.008)
Eff. range	500	**199.78 (132.05, 361.01)**	420.83 (244.9, 922.65)	414.43 (262.42, 838.88)	664.38 (448.96, 1141.24)	581.55 (383.99, 1077.52)
σ_z^2	10	–	**4.61 (2.45, 8.41)**	6.53 (3.57, 13.82)	9.32 (6.11, 14.37)	8.17 (5.85, 13.52)
ϕ_z	0.006	–	0.008 (0.003, 0.015)	0.009 (0.003, 0.021)	0.004 (0.003, 0.007)	0.006 (0.004, 0.009)
Eff. range$_z$	500	–	358.39 (198.2, 982.77)	319.25 (143.69, 940.87)	716.77 (440.01, 1138.01)	480.18 (338.1, 809.43)

J Am Stat Assoc. Author manuscript; available in PMC 2015 June 29.

Table 2

Credible intervals, 50 (2.5, 97.5) percentiles, for the synthetic analysis holdout set posterior distributions of RMSE and bias

Knots	RMSE	Bias
Direct (bias-adjusted)		
49	20.28 (19.39, 21.19)	−0.32 (−1.61, 0.91)
100	20.00 (19.13, 20.91)	−0.32 (−1.59, 0.95)
144	19.78 (18.93, 20.71)	−0.46 (−1.73, 0.82)
Hierarchical (non bias-adjusted)		
49	13.64 (12.41, 14.86)	0.15 (−0.67, 0.97)
100	13.40 (12.21, 14.68)	0.18 (−0.65, 0.99)
144	13.37 (12.11, 14.56)	0.24 (−0.52, 1.01)
Hierarchical (bias-adjusted)		
49	12.62 (11.49, 13.66)	−0.02 (−0.73, 0.67)
100	12.53 (11.33, 13.64)	−0.15 (−0.85, 0.56)
144	12.10 (10.97, 13.12)	−0.22 (−0.90, 0.47)

Table 3

Credible intervals, 50 (2.5, 97.5) percentiles in thousands, for the sum of the synthetic analysis holdout set posterior predictive distributions. The true holdout set total is $y_{\mathscr{S}_0} = 22.6 \times 10^3$

Knots	Direct	Hierarchical	Two step [$z(s_{0i})$'s known]	Two step [$z(s_{0i})$'s unknown]
49	18.9 (18.1, 19.8)	22.4 (21.6, 23.3)	22.7 (22.5, 22.8)	20.3 (19.9, 20.6)
100	19.1 (18.2, 19.9)	22.4 (21.6, 23.2)	22.7 (22.6, 22.8)	20.8 (20.2, 21.3)
144	19.2 (18.4, 20.1)	22.3 (21.6, 23.1)	22.6 (22.5, 22.7)	21.0 (20.4, 21.5)

Table 4

Parameter credible intervals, 50 (2.5 97.5) percentiles, for forest biomass nonspatial and predictive process candidate model. Note, that for brevity, only the diagonal elements of $\mathbf{LL}'$ and $\mathbf{L}_z\mathbf{L}_z'$ are provided for the space-varying coefficients model. All ϕ parameter values are scaled by 10^4

		Spatially varying intercept		Spatially varying coefficients			
Parameter	**Non spatial**	**100**	**200**	**300**	**100**	**200**	**300**
β_0	14.60 (14.34, 14.83)	14.68 (14.52, 14.83)	14.57 (14.34, 14.80)	14.63 (14.47, 14.79)	14.54 (14.35, 14.75)	14.45 (14.27, 14.64)	14.58 (14.35, 14.80)
β_{NDVI}	2.71 (2.39, 3.04)	2.72 (2.55, 2.88)	2.75 (2.33, 2.95)	2.78 (2.43, 2.96)	2.66 (2.48, 2.90)	2.89 (2.68, 3.17)	2.64 (2.14, 2.92)
β_{PRECIP}	0.25 (0.09, 0.42)	0.16 (−0.01, 0.35)	0.04 (−0.19, 0.34)	0.28 (0.17, 0.50)	−0.18 (−0.42, 0.02)	0.08 (−0.09, 0.25)	0.37 (0.20, 0.53)
β_{TMAX}	−0.08 (−0.33, 0.17)	−0.13 (−0.32, 0.30)	−0.13 (−0.40, 0.14)	−0.08 (−0.28, 0.12)	0.22 (0.01, 0.46)	0.06 (−0.29, 0.17)	−0.09 (−0.35, 0.12)
β_{TMIN}	0.31 (0.07, 0.56)	0.36 (0.05, 0.61)	0.41 (0.25, 0.63)	0.28 (0.12, 0.45)	0.10 (−0.09, 0.27)	0.38 (0.21, 0.66)	0.49 (0.28, 0.72)
η_0	−2.37 (−2.51, −2.23)	−2.76 (−2.90, −2.39)	−2.58 (−2.91, −2.40)	−2.68 (−2.83, −2.58)	−2.46 (−2.61, −2.37)	−2.45 (−2.63, −2.34)	−2.38 (−2.51, −2.26)
η_{TC1}	−5.46 (−5.73, −5.18)	−5.43 (−5.62, −5.02)	−5.29 (−5.59, −5.05)	−5.61 (−5.82, −5.33)	−5.45 (−5.70, −5.19)	−5.28 (−5.56, −5.07)	−5.23 (−5.46, −4.93)
η_{TC2}	4.54 (4.24, 4.83)	4.59 (4.14, 4.78)	4.47 (4.23, 4.80)	4.79 (4.56, 5.04)	4.65 (4.20, 4.88)	4.63 (4.37, 4.90)	4.48 (4.10, 4.68)
η_{TC3}	−1.43 (−1.65, −1.20)	−1.23 (−1.39, −1.01)	−1.17 (−1.42, −0.97)	−1.34 (−1.65, −1.14)	−1.47 (−1.70, −1.17)	−1.42 (−1.64, −1.17)	−1.36 (−1.55, −1.10)
τ^2	13.54 (12.91, 14.24)	13.21 (12.56, 13.88)	13.25 (12.66, 13.82)	13.23 (12.62, 13.86)	13.25 (12.65, 14.12)	13.32 (12.59, 13.97)	13.28 (12.71, 13.90)
$\mathbf{L}^{0,0}$	–	1.61 (1.37, 1.86)	1.77 (1.26, 2.03)	1.35 (1.18, 1.49)	0.42 (0.26, 0.65)	0.25 (0.21, 0.42)	0.45 (0.36, 0.57)
$\mathbf{LL}'_{1,1}$	–	–	–	–	0.20 (0.17, 0.34)	0.42 (0.28, 0.51)	0.55 (0.46, 0.66)
$\mathbf{LL}'_{2,2}$	–	–	–	–	0.31 (0.24, 0.45)	0.35 (0.20, 0.50)	0.47 (0.29, 0.67)
$\mathbf{LL}'_{3,3}$	–	–	–	–	0.33 (0.24, 0.46)	0.28 (0.22, 0.42)	0.42 (0.36, 0.48)
$\mathbf{LL}'_{4,4}$	–	–	–	–	0.51 (0.39, 0.65)	0.50 (0.43, 0.59)	0.33 (0.26, 0.47)
$\mathbf{LL}'_{z;0,0}$	–	0.62 (0.61, 0.64)	0.63 (0.61, 0.65)	0.61 (0.60, 0.63)	0.57 (0.48, 0.68)	0.42 (0.34, 0.59)	0.72 (0.59, 0.88)
$\mathbf{LL}'_{z;1,1}$	–	–	–	–	0.69 (0.53, 0.84)	0.48 (0.33, 0.61)	0.53 (0.46, 0.65)
$\mathbf{LL}'_{z;2,2}$	–	–	–	–	0.49 (0.34, 0.58)	0.94 (0.69, 1.37)	0.50 (0.43, 0.64)
$\mathbf{LL}'_{z;3,3}$	–	–	–	–	0.82 (0.61, 1.18)	0.99 (0.82, 1.16)	0.92 (0.74, 1.28)
ϕ_0	–	0.09 (0.07, 0.12)	0.15 (0.07, 0.18)	0.11 (0.09, 0.15)	0.08 (0.06, 0.11)	0.13 (0.10, 0.15)	0.09 (0.07, 0.10)

		Spatially varying intercept		Spatially varying coefficients			
Parameter	**Non spatial**	**100**	**200**	**300**	**100**	**200**	**300**
ϕ_1	–	–	–	–	0.17 (0.14, 0.20)	0.10 (0.08, 0.13)	0.11 (0.09, 0.12)
ϕ_2	–	–	–	–	0.09 (0.07, 0.11)	0.13 (0.09, 0.16)	0.17 (0.13, 0.22)
ϕ_3	–	–	–	–	0.11 (0.09, 0.12)	0.10 (0.09, 0.11)	0.09 (0.08, 0.10)
ϕ_4	–	–	–	–	0.08 (0.07, 0.09)	0.15 (0.12, 0.21)	0.10 (0.09, 0.11)
$\phi_{z,0}$	–	0.14 (0.10, 0.19)	0.11 (0.08, 0.15)	0.12 (0.09, 0.17)	0.12 (0.11, 0.13)	0.11 (0.09, 0.13)	0.08 (0.07, 0.09)
$\phi_{z,1}$	–	–	–	–	0.09 (0.07, 0.11)	0.10 (0.09, 0.12)	0.11 (0.10, 0.15)
$\phi_{z,2}$	–	–	–	–	0.06 (0.05, 0.06)	0.09 (0.07, 0.11)	0.08 (0.07, 0.10)
$\phi_{z,3}$	–	–	–	–	0.07 (0.06, 0.08)	0.06 (0.05, 0.06)	0.08 (0.06, 0.08)

Table 5

Credible intervals, 50 (2.5 97.5) percentiles, for the forest biomass analysis holdout set posterior distributions of deviance and RMSE scoring rules

Model	−2 log deviance	RMSE
Non spatial	−3796.52 (−3799.65, −3792.70)	7.12 (6.72, 7.60)
Spatially-varying intercept		
100	−3818.42 (−3828.59, −3812.10)	7.02 (6.60, 7.49)
200	−3835.99 (−3842.58, −3827.68)	6.92 (6.39, 7.35)
300	−3853.66 (−3866.28, −3827.11)	6.81 (6.29, 7.41)
Spatially-varying coefficients		
100	−3880.12 (−3888.99, −3869.60)	6.43 (5.99, 6.89)
200	−3870.06 (−3879.71, −3861.68)	6.59 (6.14, 7.04)
300	−3877.81 (−3886.95, −3866.36)	6.43 (6.00, 6.85)