

BAYESIAN HIERARCHICAL SPATIAL MODELS TO IMPROVE FOREST VARIABLE  
PREDICTION AND MAPPING WITH LIGHT DETECTION AND RANGING DATA  
SETS

By

Chad Babcock

A THESIS

Submitted  
to Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Geography – Master of Science

2014

## ABSTRACT

### BAYESIAN HIERARCHICAL SPATIAL MODELS TO IMPROVE FOREST VARIABLE PREDICTION AND MAPPING WITH LIGHT DETECTION AND RANGING DATA SETS

By

Chad Babcock

Light Detection and Ranging (LiDAR) data has shown great potential to estimate spatially explicit forest variables, including above-ground biomass, stem density, tree height, and more. Due to its ability to garner information about the vertical and horizontal structure of forest canopies effectively and efficiently, LiDAR sensors have played a key role in the development of operational air and space-borne instruments capable of gathering information about forest structure at regional, continental, and global scales. Combining LiDAR datasets with field-based validation measurements to build predictive models is becoming an attractive solution to the problem of quantifying and mapping forest structure for private forest land owners and local, state, and federal government entities alike. As with any statistical model using spatially indexed data, the potential to violate modeling assumptions resulting from spatial correlation is high. This thesis explores several different modeling frameworks that aim to accommodate correlation structures within model residuals. The development is motivated using LiDAR and forest inventory datasets. Special attention is paid to estimation and propagation of parameter and model uncertainty through to prediction units. Inference follows a Bayesian statistical paradigm. Results suggest the proposed frameworks help ensure model assumptions are met and prediction performance can be improved by pursuing spatially enabled models.

## ACKNOWLEDGMENTS

I would like to thank my adviser Dr. Andrew Finley for his academic guidance over the past four years. I'd also like to acknowledge my committee members Dr. Ashton Shortridge and Dr. David Macfarlane for their invaluable comments and criticisms of my thesis work. Hopefully their inputs can help to turn the following thesis chapters to peer reviewed publications.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	<b>vi</b>
<b>LIST OF FIGURES</b> . . . . .	<b>vii</b>
<b>CHAPTER 1 OVERVIEW</b> . . . . .	<b>1</b>
BIBLIOGRAPHY . . . . .	4
<b>CHAPTER 2 MULTIVARIATE SPATIAL REGRESSION MODELS FOR PREDICTING INDIVIDUAL TREE STRUCTURE VARI- ABLES USING LIDAR DATA</b> . . . . .	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Materials . . . . .	9
2.2.1 Study area and field data . . . . .	9
2.2.2 LiDAR data . . . . .	10
2.3 Methods . . . . .	11
2.3.1 Univariate spatial regression . . . . .	11
2.3.1.1 Prediction . . . . .	13
2.3.2 Multivariate spatial regression . . . . .	14
2.3.3 Model selection . . . . .	17
2.4 Candidate models and implementation . . . . .	19
2.5 Results and Discussion . . . . .	20
2.6 Summary . . . . .	26
2.7 Acknowledgment . . . . .	28
BIBLIOGRAPHY . . . . .	29
<b>CHAPTER 3 IMPROVING LIDAR BASED PREDICTION OF FOR- EST BIOMASS USING HIERARCHICAL MODELS WITH SPATIALLY VARYING COEFFICIENTS</b> . . . . .	<b>33</b>
3.1 Introduction . . . . .	33
3.2 Study Sites . . . . .	35
3.2.1 Fraser Experimental Forest . . . . .	35
3.2.2 Marcell Experimental Forest . . . . .	36
3.2.3 Glacier Lake Ecosystem Experimental Site . . . . .	36
3.2.4 Niwot Long Term Ecological Research Site . . . . .	37
3.2.5 Field Data Preparation . . . . .	37
3.2.6 LiDAR Data Preparation . . . . .	38
3.3 Methods . . . . .	42
3.3.1 Modeling Framework . . . . .	42
3.3.2 Candidate Models . . . . .	45

3.3.3	Fit and Prediction Accuracy Assessment . . . . .	46
3.4	Results . . . . .	47
3.4.1	FEF Results . . . . .	52
3.4.2	MEF Results . . . . .	57
3.4.3	GLEES Results . . . . .	59
3.4.4	NIWOT Results . . . . .	61
3.5	Discussion . . . . .	65
	BIBLIOGRAPHY . . . . .	67
<b>CHAPTER 4 SUMMARY AND FUTURE WORK . . . . .</b>		<b>71</b>

## LIST OF TABLES

Table 2.1	Non-spatial multivariate regression model parameter estimates. . . . .	23
Table 2.2	Spatial multivariate regression model parameter estimates. . . . .	24
Table 2.3	Candidate models' fit and predictive performance. . . . .	25
Table 3.1	Candidate model parameter estimates and 95% credible intervals for FEF.	53
Table 3.2	Candidate model parameter estimates and 95% credible intervals for MEF.	58
Table 3.3	Candidate model parameter estimates and 95% credible intervals for GLEES.	59
Table 3.4	Candidate model parameter estimates and 95% credible intervals for NIWOT.	62

## LIST OF FIGURES

Figure 2.1	Location of the Penobscot Experimental Forest (PEF), Maine, is identified with a diamond symbol in the lower-left inset map. Point symbols within the PEF’s two bounding polygons denote the location of the 19 inventory plots included in the study. Three of these plots with associated tree locations are illustrated in Figure 2.2. . . . .	11
Figure 2.2	Three of the 19 inventory plots with observed trees denoted with an open circle symbol and validation holdout trees denoted with a plus symbol.	18
Figure 2.3	Histogram of the spatial multivariate regression model residuals for height.	22
Figure 2.4	Scatter plots of residuals from non-spatial version of the multivariate candidate model. . . . .	26
Figure 3.1	Aerial photographs of FEF, MEF, GLEES, and NIWOT with inventory sub-plot locations highlighted in red. . . . .	38
Figure 3.2	The cumulative sums of the standardized eigenvalues (in decreasing order) for FEF, MEF, GLEES, and NIWOT. . . . .	41
Figure 3.3	Semivariogram models generated using the residuals from the <i>non-spatial</i> model estimates for FEF, MEF, GLEES, and NIWOT. . . . .	49
Figure 3.4	Maps showing predicted AGB means for the full LiDAR extent at FEF, MEF, GLEES, and NIWOT. . . . .	50
Figure 3.5	Maps showing predicted AGB 95% credible interval width (a measure of prediction uncertainty) at FEF, MEF, GLEES, and NIWOT. . . . .	51
Figure 3.6	Plots showing fitted versus actual AGB values for FEF’s <i>non-spatial</i> and <i>SVC</i> models. Vertical grey error bars depict 95% credible intervals for fitted values. . . . .	54
Figure 3.7	Maps showing the spatially varying coefficients ( $\hat{\beta}_x + \hat{w}_x$ ) and associated 95% credible interval width (CIW) for the <i>SVC</i> model at FEF. . . . .	55

Figure 3.8	Maps showing the spatially varying coefficients $\hat{\beta}_0 + \hat{w}_0$ and $\hat{\beta}_2 + \hat{w}_2$ zoomed into the extent outlined in red in figures 3.7a and 3.7c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green. . . . .	56
Figure 3.9	Plots showing fitted versus actual AGB values for MEF's <i>non-spatial</i> and <i>SVI</i> models. Vertical grey error bars depict 95% credible intervals for fitted values. . . . .	57
Figure 3.10	Plots showing fitted versus actual AGB values for GLEES's <i>non-spatial</i> and <i>SVI</i> models. Vertical grey error bars depict 95% credible intervals for fitted values. . . . .	60
Figure 3.11	Plots showing fitted versus actual AGB values for NIWOT's <i>non-spatial</i> and <i>SVC-<math>\beta_2</math></i> models. Vertical grey error bars depict 95% credible intervals for fitted values. . . . .	62
Figure 3.12	Maps showing the spatially varying coefficients ( $\hat{\beta}_x + \hat{w}_x$ ) and associated 95% credible interval width (CIW) for the <i>SVC-<math>\beta_2</math></i> model at NIWOT. . .	63
Figure 3.13	Maps showing the spatially varying coefficients $\hat{\beta}_0 + \hat{w}_0$ and $\hat{\beta}_2 + \hat{w}_2$ zoomed into the extent outlined in red in figures 3.12a and 3.12c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green. . . . .	64

# CHAPTER 1

## OVERVIEW

Over the past decade, there has been increasing interest in deriving spatially explicit estimates of forest structure variables ranging from the individual tree to landscape level (Brunner (1998); Frescino et al. (2001); Vepakomma et al. (2008); Honey-Rosés et al. (2011)).

From a forest ecology and management perspective, researchers are often interested in the spatial configuration of tree canopies, for either studying light penetration to the forest floor, animal habitat fragmentation, or other tree canopy dependent phenomena. For example, deGravelles et al. (2014) looked at Baldcypress sapling growth responses to over-story canopy gap size. Garabedian et al. (2014) used LiDAR predicted forest attributes to map and analyze red-cockaded woodpecker habitat in South Carolina. Yang et al. (2013) used a LiDAR reconstructed forest and thermal images to better understand bat flight behavior.

Beyond exploration of small scale questions concerning individual trees or stands, there is a need within the research community for statistical models appropriate for capturing spatial and temporal processes at broad scales. For example, concerning climate change, the United Nations has set forth their Reducing Emissions from Deforestation and forest Degradation (REDD+) initiative which aims to reduce net greenhouse gas emissions, in part, by encouraging the use of enhanced forest management practices in developing countries. In order to implement some of the incentive programs proposed in REDD+, methods to accurately measure and map forest carbon stocks along with associated measures of uncertainty need to be developed. The National Aeronautics and Space Administration (NASA) has created the Carbon Monitoring System (CMS) to fund research activities leading to the development of carbon measurement and verification systems in effort to meet the goals set forth by REDD+.

Calibrating LiDAR data with field validation measurements through linear regression models is often used to accommodate the need for spatial prediction of forest structure (Means et al. (2000); Montesano et al. (2014); Næsset et al. (2013)). Census collection of spatially explicit tree data is prohibitively expensive even for small forest areas. As a result, tree measurements are usually only collected for a subset of the trees in a domain. Given that most stems, even in relatively small study areas, e.g. 10 ha, are never measured, effective methods are needed to predict forest characteristics at unmeasured locations. Because LiDAR remote sensing is an effective and efficient way to collect spatially explicit data about forest height and canopy size, relating variables derived from LiDAR point clouds or waveforms to spatially indexed field measurements can lead to powerful predictive models for mapping forest structure (Wulder et al. (2008); Lefsky et al. (2002)). Special considerations are needed when working with spatial data in linear modeling. Often when developing these types of models, after accounting for LiDAR covariates, extraneous spatially structured variation is present in the residuals. This is a violation of the assumption of independent and identically distributed errors, which calls into question any inference gained or predictions carried out using the model.

This thesis details and illustrates Bayesian hierarchical spatial models appropriate for addressing issues of spatially correlated regression model residuals when calibrating LiDAR data in a forest setting. Chapter 2 explores the use of multivariate spatial regression modeling to simultaneously predict tree height, canopy depth, crown radius and stem diameter. Spatial random effects are used to ensure statistical validity and improve prediction performance. Chapter 3 also uses spatial random effects to derive spatially varying coefficients models to predict forest above-ground biomass at four study sites. Results from both Chapters 2 and 3 show that leveraging residual spatial dependence via spatial random effects can improve predictions of forest structure variables. We see from Chapter 2 that when predicting multiple, correlated, forest outcome variables, ‘borrowing strength’ both within and among

trees via random effects can lead to improved model fit and prediction. Chapter 3 shows that accommodating LiDAR coefficient non-stationarity with strategically placed spatial random effects can improve prediction accuracies. Such methods may enjoy expanded use as researchers are increasingly focusing on modeling small scale environmental processes over larger domains.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Brunner, A. (1998). A light model for spatially explicit forest stand models. *Forest Ecology and Management*, 107, 19 – 46.
- deGravelles, W.W., Hutchinson, S., & Conner, W.H. (2014). Survival and growth of suppressed baldcypress reproduction in response to canopy gap creation in a north carolina, usa swamp. *Wetlands*, 34, 79–87.
- Frescino, T.S., Edwards, T.C., & Moisen, G.G. (2001). Modeling spatially explicit forest structural attributes using generalized additive models. *Journal of Vegetation Science*, 12, 15–26.
- Garabedian, J.E., McGaughey, R.J., Reutebuch, S.E., Parresol, B.R., Kilgo, J.C., Moorman, C.E., & Peterson, M.N. (2014). Quantitative analysis of woodpecker habitat using high-resolution airborne lidar estimates of forest structure and composition. *Remote Sensing of Environment*, 145, 68 – 80.
- Honey-Rosés, J., Baylis, K., & Ramirez, M.I. (2011). A spatially explicit estimate of avoided forest loss. *Conservation Biology*, 25, 1032–1043.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 52, 19–30.
- Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., & Hendrix, C.J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66, 1367–1371.
- Montesano, P., Nelson, R.F., Dubayah, R.O., Sun, G., Cook, B.D., Ranson, K.J.R., Næsset, E., & Kharuk, V. (2014). The uncertainty of biomass estimates from lidar and sar across a boreal forest structure gradient. *Remote Sensing of Environment*, , –.
- Næsset, E., Bollandås, O.M., Gobakken, T., Gregoire, T.G., & Ståhl, G. (2013). Model-assisted estimation of change in forest biomass over an 11 year period in a sample survey supported by airborne lidar: A case study with post-stratification to provide “activity data”. *Remote Sensing of Environment*, 128, 299 – 314.
- Vepakomma, U., St-Onge, B., & Kneeshaw, D. (2008). Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. *Remote Sensing of Environment*, 112, 2326 – 2340.

- Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., & White, J.C. (2008). The role of lidar in sustainable forest management. *The Forestry Chronicle*, 84, 807–826.
- Yang, X., Schaaf, C., Strahler, A., Kunz, T., Fuller, N., Betke, M., Wu, Z., Wang, Z., Theriault, D., Culvenor, D., Jupp, D., Newnham, G., & Lovell, J. (2013). Study of bat flight behavior by combining thermal image analysis with a lidar forest reconstruction. *Canadian Journal of Remote Sensing*, 39, S112–S125.

## CHAPTER 2

# MULTIVARIATE SPATIAL REGRESSION MODELS FOR PREDICTING INDIVIDUAL TREE STRUCTURE VARIABLES USING LIDAR DATA

### 2.1 Introduction

Recent advances in remote sensing, specifically Light Detection and Ranging (LiDAR) sensors, provide detailed data at unprecedented scales. At broad spatial scales, large amounts of LiDAR data for temperate North American forests are being collected, or will soon be collected by the U.S. Forest Service (USFS), NASA's Laser Vegetation Imaging Sensor (LVIS; <https://lvis.gsfc.nasa.gov>), National Ecological Observatory Network (NEON) Airborne Observation Platform (Kamper et al., 2010), and upcoming NASA missions. These high-dimensional data contain information about individual tree and forest structure that can be related to variables of interest through suitable modeling frameworks. Many studies couple LiDAR variables with georeferenced forest inventory plot data through parametric or non-parametric regression techniques, see, e.g., Salas et al. (2010), Dalponte et al. (2009), Dinuls et al. (2012), Monnet et al. (2011), Niemann et al. (2012), Junttila et al. (2010), and references therein. The inference garnered through these analyses often supports decisions with important economic and ecological implications; therefore, it is critical to correctly estimate uncertainty.

NEON and similar initiatives aim to assimilate high spatial resolution LiDAR data with intensive forest inventories, e.g., stem maps, to better predict individual tree characteristics (Kamper et al., 2010). Given variable measurements on a subset of trees included in a stem map and LiDAR variables for all trees, a regression model is used to predict variable values for all unmeasured trees. Many studies have used regression analysis in this capacity; however, few explicitly accommodate residual spatial dependence, see, e.g., Anderson

et al. (2008), Gonzalez et al. (2010), Muss et al. (2011), Tonolli et al. (2011), Popescu et al. (2011). These non-spatial regression models are adequate in the absence of extraneous structured variation, beyond what is explained by the covariates. However, when observations are spatially indexed, we might expect similar outcomes, i.e., tree measurements, in proximate locations, possibly resulting from common environmental conditions or disturbance regimes. Ignoring this spatial dependence can result in spurious estimates of model parameters and erroneous predictions. Hoeting (2009) offers a nice discussion on the consequences of not meeting the assumption of uncorrelated model residuals. A common solution to spatial dependence among the residuals is to add a spatially varying model intercept via spatial random effects that account for spatial association through a decreasing function of distance and perhaps direction between observed locations. Beyond helping to ensure the statistical validity of the model, the addition of spatial random effects to the intercept allows for partitioning of residual uncertainty into a spatial and non-spatial component which can improve model fit and predictive performance. Further, from a forest assessment standpoint, we are often interested in predicting multiple tree variables. These variables are often correlated within and between individual trees and therefore it is attractive to consider joint models that are capable of estimating these covariances and leverage them for subsequent prediction for unmeasured trees.

In this paper, we detail spatial regression models that incorporate information from LiDAR along with information inherent in the observed forest inventory data to predict multiple outcome variables. In particular, we define and assess the utility of a multivariate spatial regression model with spatial random effects that accommodate dependence of outcome variables both within and between tree locations. The remainder of the paper is organized as follows. The study area and data set used to illustrate the proposed models are described in Section 2.2. The proposed regression models are detailed in Section 3.3 along with the associated model validation methods. Details about candidate model implementa-

tion and subsequent analysis results are given in Sections 2.4 and 3.4, respectively. Finally, some concluding remarks with an indication of future work are provided in Section 3.5.

## 2.2 Materials

### 2.2.1 Study area and field data

This study was conducted on the 1619 ha Penobscot Experimental Forest (PEF) located in the towns of Bradley and Eddington, Maine (44° 52' N, 68° 38' W; Figure 1). Mean annual temperature is 6.2°C and mean annual precipitation is 110 cm. The principal soil material is glacial till and soils range from well-drained loams and sandy loams found on glacial till ridges to poorly and very poorly drained silt loams found in flatter areas between ridges (Sendak et al., 2003).

Located within the Acadian forest, the PEF is characterized by a mixture of northern conifer and hardwood species. Conifer species include red (*Picea rubens* Sarg.), white (*P. glauca* (Moench) Voss), black spruce (*P. mariana* (Mill.) BSP), balsam fir (*Abies balsamea* (L.) Mill.), eastern hemlock (*Tsuga canadensis* (L.) Carr.), eastern white pine (*Pinus strobus* L.), and northern white-cedar (*Thuja occidentalis* L.). Hardwood species include red maple (*Acer rubrum* L.), paper (*Betula papyrifera* Marsh.), gray birch (*B. populifolia* Marsh.), quaking (*Populus tremuloides* Michx.), and bigtooth aspen (*P. grandidentata* Michx.). An operational-scale experiment to compare ten silvicultural treatments was established by USFS researchers on the PEF between 1952 and 1957. The experiment was designed to investigate the influence of silviculture on the growth, yield, and economics of eastern spruce-fir (northern conifer) stands.

A network of permanent sample plots (PSPs) was established along transects nested within each experimental unit, averaging 18 plots per unit. The PSPs consisted of a nested design with 0.081-, 0.020-, and 0.008 ha circular plots sharing the same plot center. The

USFS routinely inventoried these PSPs: all pole- and sawtimber-sized trees  $\geq 11.4$  cm diameter at breast height (DBH; measured at 1.37 m above the forest floor) were measured in the 0.081 ha plot, large saplings ( $6.4 \text{ cm} \leq \text{DBH} < 11.4 \text{ cm}$ ) were measured on the 0.020 ha plot, and small saplings ( $1.3 \text{ cm} \leq \text{DBH} < 6.4 \text{ cm}$ ) were measured on the 0.008 ha plot. Since 2000, PSPs have been remeasured by the USFS every 10 years, and, if harvesting occurred, immediately pre- and post-harvest. Live trees were uniquely numbered in these plots while experimental unit, plot and tree number, species, DBH, and status (e.g. live tree versus snag), were recorded. On approximately 25% of the PSPs, all trees were stem mapped and measured for total height (HT; nearest 0.1 m), height to crown base (HCB; lowest live branch; nearest 0.1 m), and crown radii (CR; nearest 0.1 m) in four cardinal directions. From this data, crown length (CL; m) was estimated as the absolute difference between HT and HCB. For this analysis, a total of 494 trees from 19 plots was used.

### **2.2.2 LiDAR data**

Discrete return LiDAR was acquired over the PEF in the early fall of 2010 (leaf-on) with a small fixed-wing aircraft. The density of the LiDAR returns was 0.5 pulse hits per  $\text{m}^2$ . LiDAR returns within a 5 m radius of each tree centroid were used to calculate a range of percentile heights to serve as a candidate set of LiDAR covariates. Percentile heights were calculated at 5% intervals, ranging from 5% to 100%. Exploratory data analysis using univariate non-spatial regression models identified the 95th percentile height as the best predictor variable for HT, CL, CR, and DBH. The 50th percentile height covariate also explained a statistically significant portion of the variability in HT and DBH and was not highly correlated with the 95th percentile height covariate. These covariates are referred to as P50 and P95 in the subsequent development.

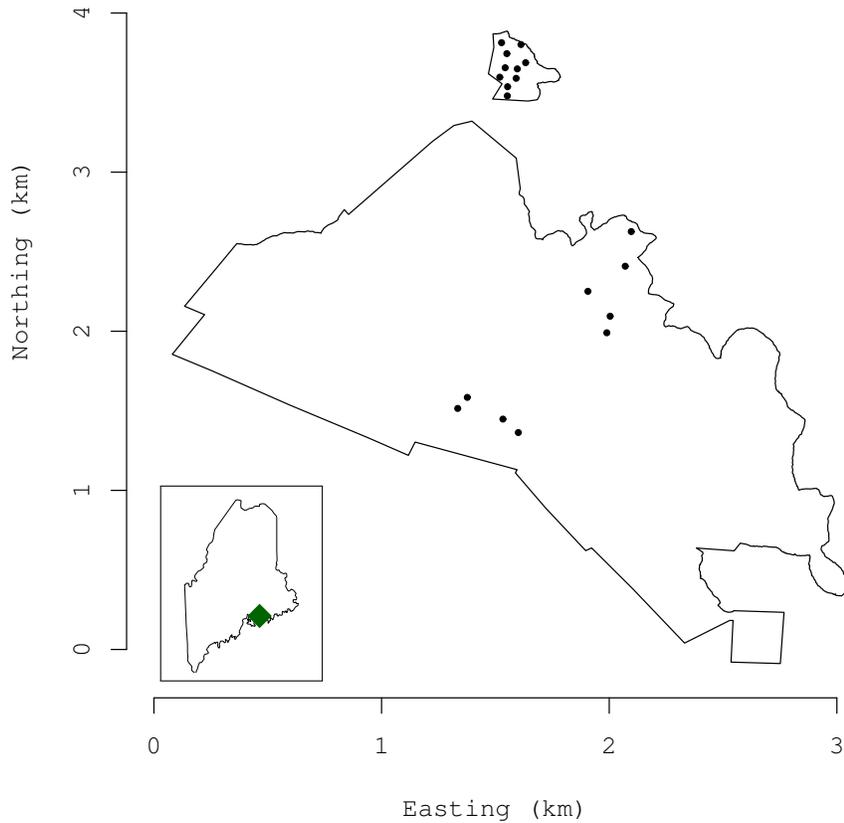


Figure 2.1: Location of the Penobscot Experimental Forest (PEF), Maine, is identified with a diamond symbol in the lower-left inset map. Point symbols within the PEF’s two bounding polygons denote the location of the 19 inventory plots included in the study. Three of these plots with associated tree locations are illustrated in Figure 2.2.

## 2.3 Methods

### 2.3.1 Univariate spatial regression

Following the data description in Section 2.2, at tree location  $\mathbf{s} \in D \subseteq \mathbb{R}^2$ , we observe an outcome variable  $y(\mathbf{s})$ , e.g., HT, CL, CR, or DBH, along with a  $p \times 1$  vector of spatially ref-

erenced covariates  $\mathbf{x}(\mathbf{s})$  derived from the LiDAR data. The outcome variable and covariates can be associated through a spatial regression model such as,

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (2.1)$$

where the model residual comprises a spatial process,  $w(\mathbf{s})$ , and an independent white-noise process,  $\epsilon(\mathbf{s})$ , that captures measurement error or micro-scale spatial variation. With any collection of  $n$  locations, say  $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ , we assume the independent and identically distributed  $\epsilon(\mathbf{s}_i)$ 's follow a normal distribution  $N(0, \tau^2)$ , where  $\tau^2$  is often called the *nugget*.

The spatial random effects,  $w(\mathbf{s})$ , provide local adjustment (with structured dependence) to the mean, interpreted as capturing the effect of unmeasured or unobserved covariates with spatial pattern. A popular modeling choice for the spatial random effects is the *Gaussian process*,  $w(\mathbf{s}) \sim GP(0, C(\cdot, \cdot; \boldsymbol{\theta}))$ , specified by a valid covariance function  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) = \text{Cov}(w(\mathbf{s}), w(\mathbf{s}^*))$  that models the covariance corresponding to a pair of locations  $\mathbf{s}$  and  $\mathbf{s}^*$ . The process realizations are collected into a  $n \times 1$  vector, say  $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))'$ , which follows a multivariate normal distribution  $MVN(\mathbf{0}, \Sigma_{\mathbf{w}})$ , where  $\Sigma_{\mathbf{w}}$  is the  $n \times n$  covariance matrix of  $\mathbf{w}$  with  $(i, j)$ th element given by  $C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ . Specification of  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$  must ensure that the resulting  $\Sigma_{\mathbf{w}}$  matrix is symmetric and positive definite. This can be done by using a class of positive definite functions that are characterized as the characteristic function of a symmetric random variable, see, e.g., Cressie (1993), Chilés & Delfiner (2008), and Banerjee et al. (2004). Here, we specify  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) = \sigma^2 \rho(\mathbf{s}, \mathbf{s}^*; \phi)$  where  $\boldsymbol{\theta} = \{\sigma^2, \phi\}$ ,  $\rho(\cdot; \phi)$  is a *correlation function*, and  $\phi$  controls the rate of correlation decay of the surface  $w(\mathbf{s})$ . Then  $\text{Var}(w(\mathbf{s})) = \sigma^2$  represents a spatial variance component in the model in (3.1). This specification assumes a *stationary* and *isotropic* process. Stationarity, in spatial modeling contexts refers to the setting where the correlation function only depends on the separation between locations (i.e., translation-invariant). Isotropy goes further and specifies the dependence through the Euclidean distance between the sites  $\mathbf{s}$  and

$\mathbf{s}^*$ , i.e.,  $\|\mathbf{s} - \mathbf{s}^*\|$  (see, e.g., Banerjee et al., 2004, p22).

To complete the Bayesian specification, we assign prior distributions to the model parameters and inference proceeds by sampling from the posterior distributions of the parameters, see, e.g., Gelman et al. (2004) for further details. Note that the spatial process induces a  $MVN(\mathbf{0}, \sigma^2 R(\phi))$  distribution on  $\mathbf{w}$ , where  $R(\phi)$  is the  $n \times n$  correlation matrix whose  $(i, j)$ th element is  $\rho(\mathbf{s}_i, \mathbf{s}_j; \phi)$ . For the remaining parameters, we assign  $\boldsymbol{\beta}$  a multivariate normal prior, say  $MVN(\boldsymbol{\mu}_\beta, \Sigma_\beta)$  and the spatial variance components  $\sigma^2$  and  $\tau^2$  are assigned inverse-Gamma (*IG*) priors. The process correlation parameter,  $\phi$ , is assigned an informative prior based on characteristics of the underlying spatial domain.

Following notation similar to that used in Gelman et al. (2004), the posterior distribution of  $p(\boldsymbol{\beta}, \mathbf{w}, \sigma^2, \tau^2, \phi | \mathbf{y})$ , where  $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))'$ , is proportional to

$$\begin{aligned} p(\phi) &\times IG(\tau^2 | a_\tau, b_\tau) \times IG(\sigma^2 | a_\sigma, b_\sigma) \\ &\times MVN(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \Sigma_\beta) \times MVN(\mathbf{w} | \mathbf{0}, \sigma^2 R(\phi)) \\ &\times \prod_{i=1}^n N(y(\mathbf{s}_i) | \mathbf{x}(\mathbf{s}_i)' \boldsymbol{\beta} + w(\mathbf{s}_i), \tau^2). \end{aligned} \quad (2.2)$$

The posterior distribution of each parameter was estimated using a Markov chain Monte Carlo (MCMC) algorithm. Specifically,  $\boldsymbol{\beta}$  and  $\mathbf{w}$  were updated using a Gibbs sampler from their full conditional distributions and the remaining parameters were updated using a Metropolis-Hastings sampler, where the target distribution for any (set of) parameter(s) is proportional to the product of the terms in (3.2) that involve that (those) parameter(s). For convenience, we collect the model parameters into  $\Omega = \{\boldsymbol{\beta}, \sigma^2, \phi, \tau^2\}$ . The MCMC algorithm yielded posterior samples  $\Omega$  and spatial random effects  $\mathbf{w}$ .

### 2.3.1.1 Prediction

As detailed in Section 3.1, our interest is in predicting structure characteristics of  $n_0$  unmeasured trees using their locations, say  $\mathcal{S}_0 = \{\mathbf{s}_{0,1}, \mathbf{s}_{0,2}, \dots, \mathbf{s}_{0,n_0}\}$ , and associated LiDAR

covariates. The posterior predictive distribution of spatial random effects  $p(\mathbf{w}_0 | \mathbf{y})$ , where  $\mathbf{w}_0 = (w(\mathbf{s}_{0,1}), w(\mathbf{s}_{0,2}), \dots, w(\mathbf{s}_{0,n_0}))'$ , is proportional to

$$\int p(\mathbf{w}_0 | \Omega, \mathbf{w}, \mathbf{y}) p(\mathbf{w} | \Omega, \mathbf{y}) p(\Omega | \mathbf{y}) d\Omega \mathbf{w}. \quad (2.3)$$

Given posterior samples,  $\{\Omega^{(l)}, \mathbf{w}^{(l)}\}_{l=1}^L$  this distribution can be obtained via composition sampling by drawing  $\mathbf{w}_0^{(l)}$  from  $p(\mathbf{w}_0 | \Omega^{(l)}, \mathbf{w}^{(l)}, \mathbf{y})$ . The process realizations over the new locations are conditionally independent of the observed outcomes given the realizations over the observed locations and the process parameters. That is,  $p(\mathbf{w}_0 | \mathbf{w}, \Omega, \mathbf{y}) = p(\mathbf{w}_0 | \Omega, \mathbf{w})$ , which is a multivariate normal distribution with mean and variance given by

$$E[\mathbf{w}_0 | \mathbf{w}, \Omega] = \text{Cov}(\mathbf{w}_0, \mathbf{w}) \text{Var}^{-1}(\mathbf{w}) \mathbf{w} = R_0(\phi)' R(\phi)^{-1} \mathbf{w}$$

and

$$\text{Var}[\mathbf{w}_0 | \mathbf{w}, \Omega] = \sigma^2 \left\{ R(\phi) - R_0(\phi)' R(\phi)^{-1} R_0(\phi) \right\},$$

where  $R_0(\phi)'$  is the  $n_0 \times n$  matrix with  $(i, j)$ th element given by  $\rho(\mathbf{s}_{0,i}, \mathbf{s}_j; \phi)$ . Finally, given a set of covariates at a new location  $\mathbf{s}_0$ , samples from the posterior predictive distribution of the outcome variable,  $y(\mathbf{s}_0)^{(l)}$ , are drawn from  $N(\mathbf{x}(\mathbf{s}_0)' \boldsymbol{\beta}^{(l)} + \mathbf{w}_0^{(l)}, \tau^2)^{(l)}$  for  $l = 1, 2, \dots, L$ .

### 2.3.2 Multivariate spatial regression

As described in Section 3.1, given the covariance among trees' structure outcome variables it is particularly attractive to specify a joint model capable of estimating and using this covariance to potentially improve prediction. Here, we extend the univariate spatial regression to consider each tree's  $m$  outcome variables. In our setting  $m = 4$ , i.e., HT, CL, CR, and DBH. Let  $\mathbf{y}(\mathbf{s}) = (y_1(\mathbf{s}), \dots, y_m(\mathbf{s}))'$  be the collection of these outcomes measured at tree location  $\mathbf{s}$ . Given such data our model should account for inherent association between the outcomes within each location and between observations across locations. We consider the

following multivariate regression model

$$\mathbf{y}(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta} + \mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}). \quad (2.4)$$

Here  $\mathbf{X}(\mathbf{s})'$  is an  $m \times p$  block-diagonal matrix ( $p = \sum_{l=1}^m p_l$ ), with  $l$ -th diagonal element being the  $1 \times p_l$  vector  $\mathbf{x}_l(\mathbf{s})'$ , and  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)'$  is a  $p \times 1$  vector of regression coefficients, with each  $\boldsymbol{\beta}_l$  a  $p_l \times 1$  vector of regression coefficients corresponding to  $\mathbf{x}_l(\mathbf{s})'$ . This specifies the mean structure that accounts for large scale variation in the outcome. Spatial variation in the outcome is modeled using a  $m \times 1$  vector of random effects  $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), \dots, w_m(\mathbf{s}))'$ . Customarily, we assume the unstructured residuals  $\boldsymbol{\epsilon}(\mathbf{s})$ , defined analogous to  $\mathbf{w}(\mathbf{s})$ , follow a multivariate normal distribution with zero mean and an  $m \times m$  dispersion matrix  $\boldsymbol{\Psi}$ .

Spatially structured dependence is introduced in (2.4) through a multivariate ( $m \times 1$ ) spatial process  $\mathbf{w}(\mathbf{s}) \sim GP(\mathbf{0}, C(\cdot, \cdot; \boldsymbol{\theta}))$ , see, e.g., Cressie (1993), where the *cross-covariance function*  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$  is now defined to be the  $m \times m$  matrix with  $(i, j)$ th entry  $C_{ij}(\mathbf{s}, \mathbf{s}^*) = \text{Cov}(w_i(\mathbf{s}), w_j(\mathbf{s}^*))$ . As in the univariate case, this model specifies a stationary and isotropic process.

The cross-covariance function completely determines the joint dispersion structure implied by the spatial process. Specifically, for any  $n$  and any arbitrary collection of locations  $\mathcal{S}$  the  $nm \times 1$  vector  $\mathbf{w} = (\mathbf{w}(\mathbf{s}_1)', \dots, \mathbf{w}(\mathbf{s}_n))'$  will have the covariance matrix given by  $\Sigma_{\mathbf{w}}(\boldsymbol{\theta})$ , an  $nm \times nm$  block matrix whose  $(i, j)$ th block is the cross-covariance matrix  $C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ , which is a symmetric and positive definite function.

One possibility for  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta} = \{\Sigma, \phi\}$ , is a separable specification,  $\rho(\mathbf{s}, \mathbf{s}^*; \phi)\Sigma$ , where  $\Sigma$  is an  $m \times m$  covariance matrix between the outcomes and  $\rho(\mathbf{s}, \mathbf{s}^*; \phi)$  is a spatial correlation function. This implies  $\text{Var}(\mathbf{w}) = R(\phi) \otimes \Sigma$ , where  $\otimes$  is the kronecker product operator and  $R(\phi)$  was defined previously in Section 2.3.1. A separable spatial association structure is rather restrictive here, assigning a single spatial correlation parameter  $\phi$  to be shared by all the outcomes. This can be inappropriate, e.g., why should all the tree structure

variables have the same strength of spatial associations?

More general 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 data be positive definite. Various constructions are possible, see, e.g., Hoef & Barry (1998), Higdon et al. (1999), and Fuentes & Smith (2002).

We motivate an alternative approach as follows. For the separable model, notice that we can write  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) = \mathbf{A}\Theta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\phi})\mathbf{A}'$  where  $\Sigma = \mathbf{A}\mathbf{A}'$  is a Cholesky factorization and  $\Theta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\phi}) = \rho(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\phi})I_m$ , where  $I_m$  is the  $m \times m$  identity matrix. We call  $\Theta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\phi})$  the *cross-correlation* function which must satisfy  $\Theta(\mathbf{s}, \mathbf{s}; \boldsymbol{\phi}) = I_m$  for all  $\boldsymbol{\phi} = \{\phi_1, \phi_2, \dots, \phi_m\}$ . This implies that  $C(\mathbf{s}, \mathbf{s}) = \mathbf{A}\mathbf{A}'$  and  $\mathbf{A}$  identifies with a matrix square-root (e.g., Cholesky) of  $C(\mathbf{s}, \mathbf{s})$ . For modeling  $\mathbf{A}$ , without loss of generality one can assume that  $\mathbf{A} = C^{1/2}(\mathbf{s}, \mathbf{s})$  is a lower-triangular square-root; the one-to-one correspondence between the elements of  $\mathbf{A}$  and  $C_{\mathbf{w}}(\mathbf{s}, \mathbf{s})$  is well-known (see, e.g., Harville, 1997, p229). Therefore,  $\mathbf{A}$  determines the association between the elements of  $\mathbf{w}(\mathbf{s})$  within  $\mathbf{s}$ . Also we adopt a flexible and computationally feasible approach that specifies  $\Theta(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\phi})$  as a diagonal matrix with  $\rho_k(\mathbf{s}, \mathbf{s}^*; \phi_k)$ ,  $k = 1, \dots, m$  as its diagonal elements. This incorporates a set of  $m$  correlation parameters, or even  $m$  different correlation functions, offering an attractive, easily interpretable and flexible approach. This approach resembles the linear model of coregionalization (LMC) as in, e.g., Grzebyk & Wackernagel (1994), Wackernagel (2003), Schmidt & Gelfand (2003), and Gelfand et al. (2004). See, also, Reich & Fuentes (2007) for a Bayesian nonparametric adaptation. We will assume this structure in our subsequent development.

Once a cross-covariance function is specified for a multivariate Gaussian process, the realizations of  $\mathbf{w}(\mathbf{s})$  over the set of observed locations  $\mathcal{S}$  is given by  $MVN(\mathbf{0}, \Sigma_{\mathbf{w}}(\boldsymbol{\theta}))$ , where  $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\phi}\}$  and  $\Sigma_{\mathbf{w}}(\boldsymbol{\theta})$  is an  $mn \times mn$  block matrix whose  $(i, j)$ th block is the  $m \times m$  cross-covariance  $C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ ,  $i, j = 1, \dots, n$ .

Collecting observations from the  $n$  locations into an  $mn \times 1$  vector  $\mathbf{y} = (\mathbf{y}(\mathbf{s}_1)', \dots, \mathbf{y}(\mathbf{s}_n)')'$ ,

the multivariate analogue of (3.2) as

$$\begin{aligned}
& \prod_{k=1}^m p(\boldsymbol{\phi}_k) \times p(\mathbf{A}) \times MVN(\boldsymbol{\beta} \mid \boldsymbol{\mu}_\beta, \Sigma_\beta) \\
& \quad \times p(\boldsymbol{\Psi}) \times MVN(\mathbf{w} \mid \mathbf{0}, \Sigma_{\mathbf{w}}(\boldsymbol{\theta})) \\
& \quad \times \prod_{i=1}^n MVN(\mathbf{y}(\mathbf{s}_i) \mid \mathbf{X}(\mathbf{s}_i)' \boldsymbol{\beta} + \mathbf{w}(\mathbf{s}_i), \boldsymbol{\Psi}), \tag{2.5}
\end{aligned}$$

With the appropriate dimensional adjustments, the prior specifications for  $\mathbf{w}$ ,  $\boldsymbol{\beta}$ , and elements of  $\boldsymbol{\phi}$  are the same as the univariate case. We assumed cross-covariance matrix  $\mathbf{A}\mathbf{A}'$  and  $\boldsymbol{\Psi}$  followed an inverse-Wishart (IW) distribution. Sampling from the parameters' posterior distributions follows the same algorithm as the univariate case. Upon convergence, the MCMC output generates  $L$  samples,  $\{\Omega^{(l)}, \mathbf{w}^{(l)}\}_{l=1}^L$ , where  $\Omega$  is now  $\{\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\phi}, \boldsymbol{\Psi}\}$ . Subsequent prediction of the spatial random effects  $\mathbf{w}(\mathbf{s}_0)$  and the outcome  $\mathbf{y}(\mathbf{s}_0)$  is also analogous to the univariate setting.

### 2.3.3 Model selection

Model fit is assessed using a posterior predictive loss function based on simulating *independent* replicates for each observed outcome. For the multivariate spatial model (2.4), and each location  $\mathbf{s}_i$  we compute, for each  $i = 1, 2, \dots, n$ ,

$$\begin{aligned}
p(\mathbf{y}_{rep}(\mathbf{s}_i) \mid \mathbf{y}) &= \int MVN(\mathbf{y}_{rep}(\mathbf{s}_i) \mid \mathbf{X}\boldsymbol{\beta} + \mathbf{w}(\mathbf{s}_i), \boldsymbol{\Psi}) \\
& \quad \times p(\Omega, \mathbf{w}(\mathbf{s}_i) \mid \mathbf{y}) d\Omega d\mathbf{w}(\mathbf{s}_i) \tag{2.6}
\end{aligned}$$

where the conditional distribution of the replicate  $\mathbf{y}_{rep}(\mathbf{s}_i)$  given the parameters is simply the likelihood component corresponding to  $\mathbf{y}(\mathbf{s}_i)$ . Models that perform well under a decision-theoretic balanced loss function, i) penalizing both departure of replicated means from their observed values (lack of fit) and ii) excessive uncertainty in the replicated data, are preferred.

Let  $\boldsymbol{\mu}_{rep,i}$  and  $\boldsymbol{\Sigma}_{rep,i}$  be the posterior predictive mean and variance for each  $\mathbf{y}_{rep}(\mathbf{s}_i)$ . Then using a squared error loss function (e.g., Gelfand & Ghosh, 1998), the measures for these two criteria are evaluated as  $G = \sum_{i=1}^n \|\mathbf{y}(\mathbf{s}_i) - \boldsymbol{\mu}_{rep,i}\|^2$ , where  $\|\cdot\|$  is the standard Euclidean norm, and  $P = \sum_{i=1}^n \text{trace}(\boldsymbol{\Sigma}_{rep,i})$ . Following Gelfand & Ghosh (1998), we used the score  $D = G + P$  as a model selection criterion, with lower values of  $D$  indicating better models.

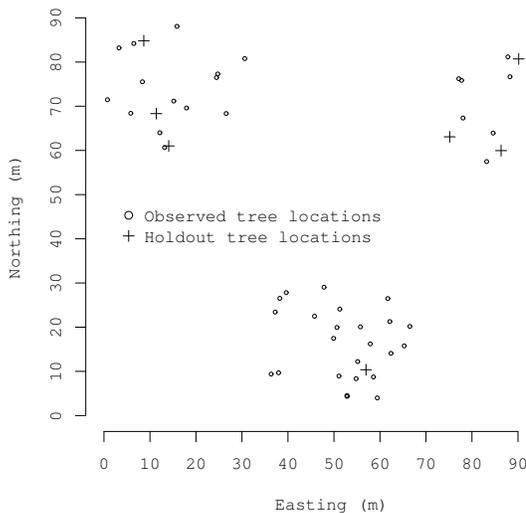


Figure 2.2: Three of the 19 inventory plots with observed trees denoted with an open circle symbol and validation holdout trees denoted with a plus symbol.

Finally, because prediction is our overarching objective, we assessed model predictive performance using a split-set validation approach. Here, 85% of the observations were used to estimate candidate models' parameters and the remaining 15% were used for subsequent prediction-based validation. Models were ranked based on their holdout set root mean squared prediction error (RMSPE), where again lower values of RMSPE indicate greater predictive performance.

## 2.4 Candidate models and implementation

We considered four candidate models for the tree structure data described in Section 2.2: *i*) univariate non-spatial regression (3.1) with  $\mathbf{w} = \mathbf{0}$ ; *ii*) univariate spatial regression (3.1); *iii*) multivariate non-spatial regression (2.4) with  $\mathbf{w} = \mathbf{0}$ , and; *iv*) multivariate spatial regression (2.4).

Prior to regression analysis HT, CL, CR, and DBH were log transformed to satisfy the model assumption that  $\epsilon(\mathbf{s}_i)$ 's are normally distributed. This transformation also ensures a correct positive support for the outcomes following back transformation.

As detailed in Section 2.3.3, model fit was evaluated using the  $D$  criterion. Predictive performance was assessed using RMSPE calculated using 74 randomly selected trees i.e., a 15% holdout set. Candidate model parameters were estimated using the remaining 420 observations. The distribution of observed and holdout trees in three inventory plots are illustrated in Figure 2.2.

To complete the proposed models' Bayesian specification, prior distributions were assigned to each parameter. As is customary, a flat prior was assigned to the regression intercept and slope parameters, i.e.,  $\Sigma_\beta = \mathbf{0}$ . The covariance matrices  $\mathbf{A}\mathbf{A}'$  and  $\Psi$ , were each assigned an  $IW(a, b)$  with the degrees of freedom,  $a$ , set to  $m + 1$ . For  $\mathbf{A}\mathbf{A}'$  and  $\Psi$ , the  $IW$  scale matrix,  $b$ , was constructed with zeros on the off-diagonal elements and diagonal elements taken as the partial-sill and nugget values, respectively, from univariate semivariograms fit to the residuals of the non-spatial multivariate model. An exponential correlation function,  $\rho(\mathbf{s}, \mathbf{s}^*; \phi) = \exp(-\phi\|\mathbf{s} - \mathbf{s}^*\|)$ , was used in the univariate and multivariate models. The maximum distance between any two trees in the data set is 2479 m. Therefore the spatial decay parameters,  $\phi$ 's, were assigned a uniform  $U(0.0012, 3)$ , which corresponds to support for an *effective spatial range* between approximately 1 to 2500 m (i.e., where effective spatial range is the distance at which the correlation drops to 0.05).

The candidate models were coded in C++ and used Intel’s Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. All analyses were conducted on a Linux workstation using two Intel Nahalem hyperthreaded quad-Xeon processors. The samplers were run for three chains of 50,000 iterations. The first 10,000 iterations were discarded as burn-in, following convergence diagnostics detailed in Gelman et al. (2004). Parameter and predictive inference was based on 120,000 post burn-in samples (i.e., three chains each of 40,000 iterations). The most computationally intensive model required approximately 3 hours to deliver the three chains.

## 2.5 Results and Discussion

Figure 2.3 offers histograms of the multivariate spatial regression model residuals and provides evidence that, following the transformation described in Section 2.4, the assumption about the distribution of residuals is reasonable. Parameter estimates for the multivariate non-spatial and spatial regression models, i.e., candidate models *iii* and *iv*, are provided in Tables 2.1 and 2.2, respectively. In both tables, subscripts on the regression slope parameters indicate the outcome variable and associated LiDAR covariate. The  $\beta_{.,0}$  correspond to the intercept whereas  $\beta_{.,P50}$  and  $\beta_{.,P95}$  correspond to the 50th and 95th LiDAR height percentile covariate, respectively. All of the non-spatial model slope parameters are significant at the 0.05 level, i.e., 95% credible intervals do not include zero. This result suggests the LiDAR covariates explain a substantial portion of the variability in the outcome variables. A very different conclusion is drawn from the multivariate spatial regression model results, Table 2.2, which suggest none of the slope parameters associated with the LiDAR covariates differ from zero. These disparate conclusions occur because we violate the key model assumption of independent and identically distributed residuals when applying the non-spatial model to these data. The non-spatial model assumes each tree observation contributes a

unique portion of information, however, this is not the case – due to strong spatial dependence among measurements between tree locations. The addition of the spatial random effects help to meet this model assumption by explicitly modeling the residual spatial dependence and hence providing a more accurate summary of covariates’ importance. Comparing Tables 2.1 and 2.2, apportioning residual variability from the non-spatial model variance-covariance  $\Psi$  to the spatial model cross-covariance  $\mathbf{A}\mathbf{A}'$  confirms the presence of strong residual spatial dependence. Further, the estimates of effective spatial range in Table 2.2 suggest there is spatial dependence beyond 0.5 km for all outcomes, even after accounting for the LiDAR covariates.

Given the assumed exponential correlation function, the effective spatial range associated with the first outcome variable in the multivariate vector, i.e.,  $\mathbf{y}_1(\mathbf{s})$ , is obtained by solving  $\rho(d; \phi) = 0.05$  for  $d$ , i.e.,  $d = -\ln(0.05)/\phi$ . However, because of the linear combination induced by the cross-covariance matrix, the subsequent effective spatial ranges are obtained by solving a system of equations (see, e.g., Gelfand et al., 2004, p292). For example, the effective spatial range for  $\mathbf{y}_2(\mathbf{s})$  is given by solving  $(a_{2,1}^2\rho(d; \phi_1) + a_{2,2}^2\rho(d; \phi_2))/(a_{2,1}^2 + a_{2,2}^2) = 0.05$  for  $d$ , where  $a_{2,1}$  and  $a_{2,2}$  are the elements of  $\mathbf{A}$  corresponding to the row and column subscripts. In a similar way, the effective spatial range for  $\mathbf{y}_3(\mathbf{s})$  is given by solving  $(a_{3,1}^2\rho(d; \phi_1) + a_{3,2}^2\rho(d; \phi_2) + a_{3,3}^2\rho(d; \phi_3))/(a_{3,1}^2 + a_{3,2}^2 + a_{3,3}^2) = 0.05$  for  $d$ . The effective spatial ranges for  $\mathbf{y}_4(\mathbf{s}), \dots, \mathbf{y}_m(\mathbf{s})$  follow the same pattern.

Although not shown, the univariate non-spatial and spatial regression models, i.e., candidate models *i* and *ii*, parameter estimates do not differ substantially from their multivariate counterparts.

Contrary to the spatial model results, one expects the P95 covariate to explain a substantial portion of variability in at least the tree height outcome, i.e., HT. The P95, or similar upper canopy LiDAR return covariate, has been shown to be well correlated with tree height in many studies. However, in the current data set there is just not enough unique

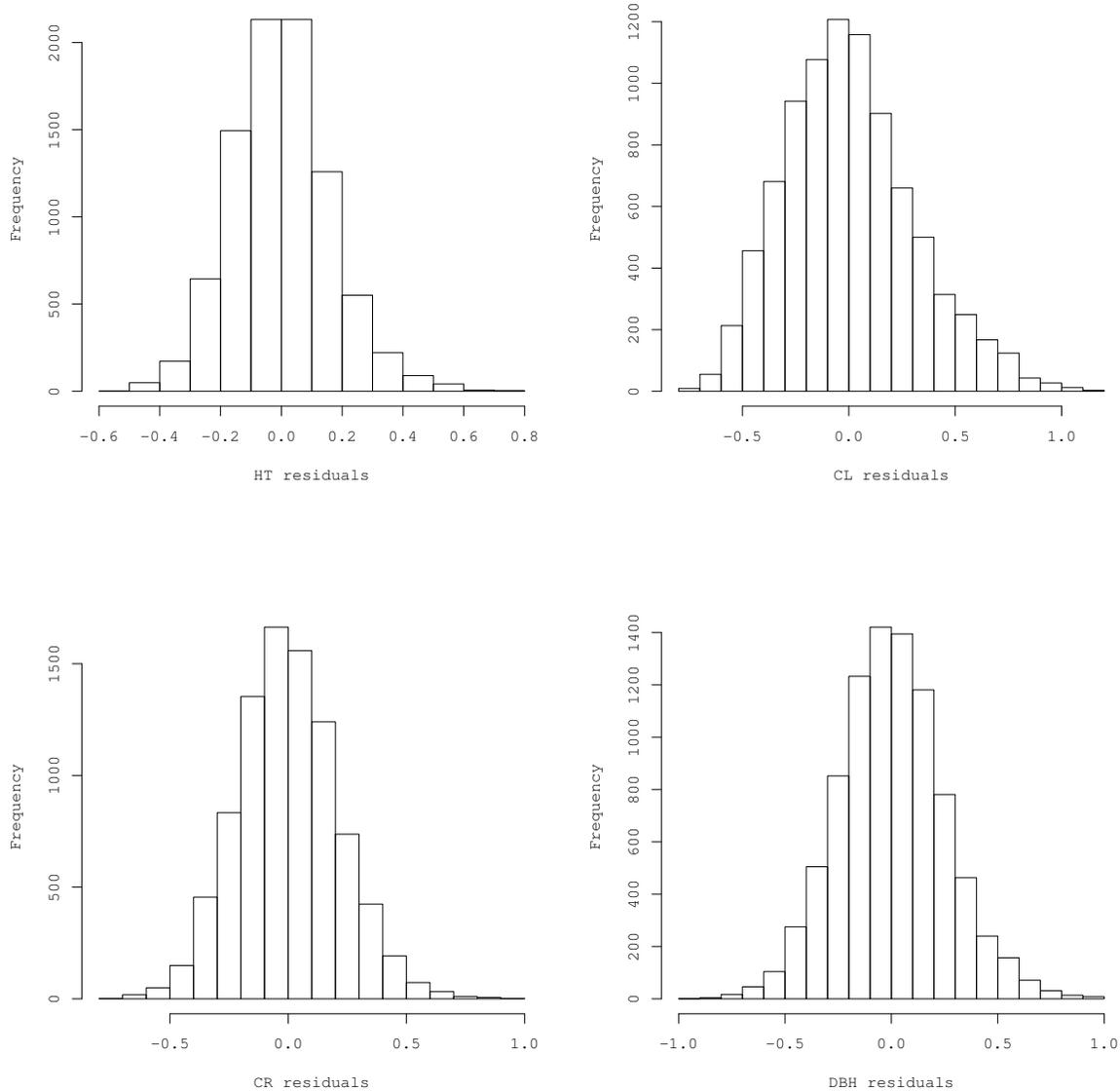


Figure 2.3: Histogram of the spatial multivariate regression model residuals for height.

information to reduce the width of  $\beta_{HT,P95}$ 's credible interval and exclude zero. Given the estimated range of spatial dependence, we might conclude that tree measurements within plots are redundant and that even plots in close proximity are not contributing much unique information. In the end, we are probably only working with an *effective* sample size of  $\lesssim 19$ ,

Table 2.1: Non-spatial multivariate regression model parameter estimates.

Parameter	50 (2.5, 97.5) percentiles
$\beta_{HT,0}$	1.67 (1.60, 1.74)
$\beta_{HT,P50}$	-0.02 (-0.03, -0.01)
$\beta_{HT,P95}$	0.07 (0.06, 0.08)
$\beta_{CL,0}$	1.29 (1.20, 1.38)
$\beta_{CL,P95}$	0.04 (0.03, 0.04)
$\beta_{CR,0}$	-0.10 (-0.20, -0.00)
$\beta_{CR,P95}$	0.05 (0.04, 0.05)
$\beta_{DBH,0}$	1.41 (1.30, 1.54)
$\beta_{DBH,P50}$	-0.01 (-0.02, -0.00)
$\beta_{DBH,P95}$	0.09 (0.08, 0.10)
$\Psi_{HT,HT}$	0.07 (0.06, 0.08)
$\Psi_{HT,CL}$	0.06 (0.05, 0.07)
$\Psi_{HT,CR}$	0.06 (0.05, 0.07)
$\Psi_{HT,DBH}$	0.10 (0.09, 0.12)
$\Psi_{CL,CL}$	0.15 (0.14, 0.17)
$\Psi_{CL,CR}$	0.11 (0.09, 0.12)
$\Psi_{CL,DBH}$	0.13 (0.11, 0.15)
$\Psi_{CR,CR}$	0.18 (0.16, 0.21)
$\Psi_{CR,CL}$	0.17 (0.14, 0.19)
$\Psi_{DBH,DBH}$	0.25 (0.22, 0.28)

as opposed to the 420 tree observations assumed when applying the non-spatial model.

Model fit and prediction summaries are given in Table 2.3. Here, a lower value of  $D$  for the spatial models compared to that of the non-spatial models suggests that the addition of spatial random effects improves model fit.

As detailed in Section 3.1, our interest is in the predictive ability of the models. This is assessed using the RMSPE based on prediction for the 15% holdout set and summarized in the last four rows of Table 2.3. These results suggest the spatially structured random effects result in greater predictive accuracy than that obtained from the non-spatial models, i.e., candidate models *ii* and *iv* have lower RMSPE than their non-spatial counterparts *i* and *iii*, respectively.

There are marginal differences between the RMSPEs obtained from the univariate and multivariate spatial models. However, with the exception of  $CL$ , the multivariate model achieves a smaller RMSPE for the outcome variables. The lower RMSPE is likely due to

Table 2.2: Spatial multivariate regression model parameter estimates.

Parameter	50 (2.5, 97.5) percentiles
$\beta_{HT,0}$	2.65 (2.36, 2.98)
$\beta_{HT,P50}$	0.00 (-0.00, 0.01)
$\beta_{HT,P95}$	0.00 (-0.01, 0.01)
$\beta_{CL,0}$	2.15 (1.69, 2.39)
$\beta_{CL,P95}$	-0.01 (-0.02, 0.01)
$\beta_{CR,0}$	0.97 (0.43, 1.24)
$\beta_{CR,P95}$	-0.00 (-0.02, 0.01)
$\beta_{DBH,0}$	3.07 (2.43, 3.54)
$\beta_{DBH,P50}$	0.00 (-0.00, 0.02)
$\beta_{DBH,P95}$	-0.00 (-0.02, 0.01)
$\mathbf{AA}_{HT,HT}$	0.14 (0.11, 0.17)
$\mathbf{AA}_{HT,CL}$	0.14 (0.11, 0.15)
$\mathbf{AA}_{HT,CR}$	0.17 (0.14, 0.18)
$\mathbf{AA}_{HT,DBH}$	0.25 (0.21, 0.27)
$\mathbf{AA}_{CL,CL}$	0.17 (0.12, 0.24)
$\mathbf{AA}_{CL,CR}$	0.19 (0.16, 0.26)
$\mathbf{AA}_{CL,DBH}$	0.29 (0.22, 0.34)
$\mathbf{AA}_{CR,CR}$	0.30 (0.24, 0.34)
$\mathbf{AA}_{CR,DBH}$	0.37 (0.31, 0.42)
$\mathbf{AA}_{DBH,DBH}$	0.51 (0.43, 0.60)
$\rho_{HT,CL}$	0.87 (0.83, 0.89)
$\rho_{HT,CR}$	0.83 (0.78, 0.85)
$\rho_{HT,DBH}$	0.91 (0.89, 0.95)
$\rho_{CL,CR}$	0.89 (0.86, 0.93)
$\rho_{CL,DBH}$	0.94 (0.92, 0.96)
$\rho_{CR,DBH}$	0.94 (0.92, 0.97)
$\Psi_{HT,HT}$	0.03 (0.02, 0.03)
$\Psi_{HT,CL}$	0.03 (0.02, 0.03)
$\Psi_{HT,CR}$	0.01 (0.01, 0.02)
$\Psi_{HT,DBH}$	0.03 (0.02, 0.04)
$\Psi_{CL,CL}$	0.09 (0.08, 0.10)
$\Psi_{CL,CR}$	0.04 (0.03, 0.05)
$\Psi_{CL,DBH}$	0.04 (0.03, 0.05)
$\Psi_{CR,CR}$	0.05 (0.04, 0.06)
$\Psi_{CR,DBH}$	0.03 (0.02, 0.04)
$\Psi_{DBH,DBH}$	0.07 (0.04, 0.09)
Eff. range <sub>1</sub> (m)	598.22 (276.58, 804.28)
Eff. range <sub>2</sub> (m)	609.64 (428.73, 740.14)
Eff. range <sub>3</sub> (m)	567.48 (352.99, 704.88)
Eff. range <sub>4</sub> (m)	592.88 (338.04, 772.92)

explicit modeling of the covariance among the random effects via the cross-covariance matrix  $\mathbf{AA}'$ . The cross-covariance matrix allows the model to *borrow strength* across the random

Table 2.3: Candidate models’ fit and predictive performance.

	Univariate		Multivariate	
	Non-spatial	Spatial	Non-spatial	Spatial
$G$	–	–	265	75
$P$	–	–	278	121
$D$	–	–	543	196
RMSPE <sub>HT</sub>	3.84	2.60	3.82	2.56
RMSPE <sub>CL</sub>	2.91	2.41	2.88	2.45
RMSPE <sub>CR</sub>	1.06	0.81	1.05	0.76
RMSPE <sub>DBH</sub>	12.67	8.67	12.73	8.46

effects which can improve prediction. Figure 2.4 shows scatter plots of the univariate non-spatial model residuals, i.e., after accounting for the LiDAR covariates. These plots suggest several strong positive linear relationships among the residuals, a portion of which is captured by the  $\mathbf{AA}'$  which can also be viewed as the *within* tree outcome covariance. In Table 2.2, we have converted the  $\mathbf{AA}'$  covariance estimates to correlations to simplify interpretation, e.g.,  $\rho_{HT,CL} = \mathbf{AA}'_{HT,CL} / (\mathbf{AA}'_{HT,HT} \times \mathbf{AA}'_{CL,CL})$ . It is these strong correlations that help improve predictive performance. Further, explicitly modeling these within tree covariances can help to ensure a biologically feasible combination of predicted structure variables for new locations – this is not guaranteed when using  $m$  separate univariate models for prediction. Finally, although it is not an objective of this study, estimates of  $\mathbf{AA}'$  or corresponding  $\rho$ ’s can allow one to test hypotheses about the statistical significance of covariation among outcomes – after accounting for covariates. For example, after accounting for the LiDAR covariates, all outcomes have positive covariance or correlations that differ significantly from zero.

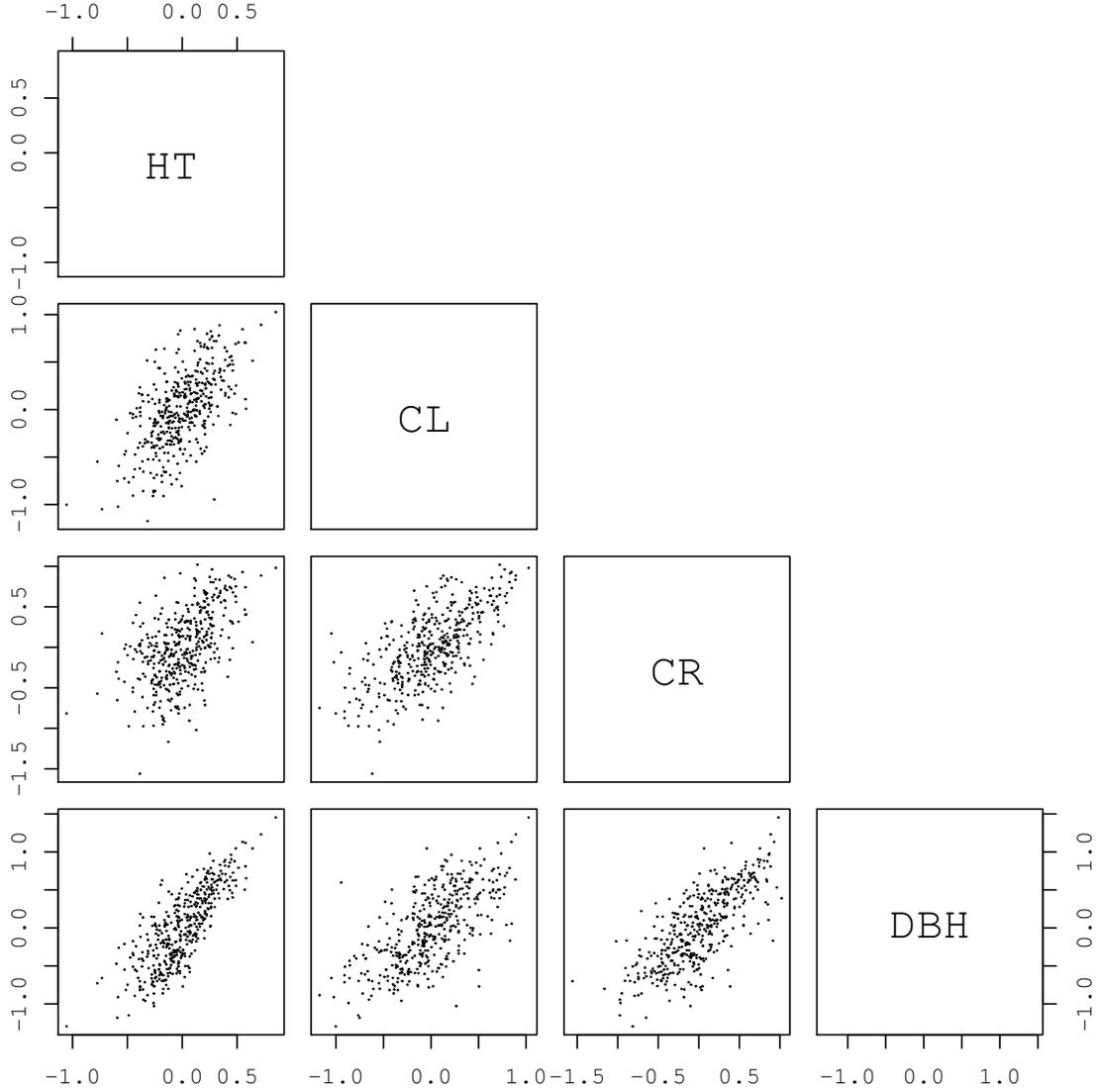


Figure 2.4: Scatter plots of residuals from non-spatial version of the multivariate candidate model.

## 2.6 Summary

In this paper we compared univariate and multivariate regression models with and without spatially-structured random effects for predicting several tree structure variables. In addition

to helping meet the regression model’s assumption of independent and identically distributed residuals, the results of the PEF data analysis suggest that the addition of univariate or multivariate spatial random effects improved model fit and prediction. In some cases, explicitly modeling the residual spatial dependence among the outcomes via multivariate random effects resulted in improved inference.

The difference between inference gleaned from the non-spatial and spatial models about the importance of the LiDAR covariates underscores the need to meet model assumptions in this and similar studies where residual spatial dependence could be present.

By working in a Bayesian paradigm we have access to the full posterior predictive distribution of each outcome variable at each new location, which facilitates uncertainty assessment. Access to posterior predictive samples could also be useful in settings where the investigator wishes to propagate uncertainty in outcome variables through subsequent economic or ecological numerical models. For example, many forest yield models require spatially and temporally explicit inputs that must be somehow imputed to the region of interest.

There are obviously many alternative approaches to deriving covariates from LiDAR data and more exhaustive approaches for selecting a subset of covariates to serve in the regression model. Regardless of the approach used, there is typically strong correlation among the covariates and hence only a few can be included in the model – without risking issues related to multicollinearity. Therefore, we do not expect that a more sophisticated covariate selection approach would result in substantial gains in explained variance over that obtained using the P50 and P95 covariates. Rather, in practice the addition of spatial random effects generally results in larger gains in model fit and predictive ability.

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

## 2.7 Acknowledgment

The work of all authors was supported by the NASA grant 11-CMS11-0021. The third author was supported by NSF EF-1137309, DMS-1106609, and NASA grant 11-CMS11-0028. This work was also supported by The National Ecological Observatory Network (NEON) which is a project sponsored by the NSF grants EF-1029808, EF-1138160, EF-1150319 and DBI-0752017 and managed under cooperative agreement by NEON, Inc. Data for this study were provided by a unit of the Northern Research Station, U.S. Forest Service, located at the Penobscot Experimental Forest in Maine. The authors would like to thank Michael R. Saunders for the use of his dissertation data.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., M. L. Smith, R. O. Dubayah, M.A.H., & Blair, B.J. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112, 1856–1870.
- Banerjee, S., Carlin, B.P., & Gelfand, A.E. (2004). Hierarchical Modeling and Analysis for Spatial Data. First ed., Chapman and Hall/CRC Press, Boca Raton, FL.
- Chilés, J.P., & Delfiner, P. (2008). Geostatistics: Modeling Spatial Uncertainty. Second ed., Wiley, New York, NY.
- Cressie, N.A.C. (1993). Statistics for Spatial Data. Second ed., Wiley, New York, NY.
- Dalponte, M., Coops, N.C., Bruzzone, L., & Gianelle, D. (2009). Analysis on the use of multiple returns lidar data for the estimation of tree stems volume. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2, 310–318.
- Dinuls, R., Erins, G., Lorencs, A., Mednieks, I., & Sinica-Sinavskis, J. (2012). Tree species identification in mixed baltic forest using lidar and multispectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5, 594–603.
- Fuentes, M., & Smith, R.L. (2002). A new class of nonstationary spatial models. *Biometrika*, 89, 197–210.
- Gelfand, A.E., & Ghosh, S.K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1–11.
- Gelfand, A.E., Schmidt, A.M., Banerjee, S., & Sirmans, C.F. (2004). Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *TEST*, 13, 263–312.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). Bayesian Data Analysis. Second ed., Chapman and Hall/CRC Press, Boca Raton, FL.
- Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., & Palace, M. (2010). Forest carbon densities and uncertainties from lidar, quickbird, and field measurements in california. *Remote Sensing of Environment*, 114, 1561–1575.
- Grzebyk, M., & Wackernagel, H. (1994). Multivariate analysis and spatial/temporal scales: Real and complex models, in: Proceedings of the XVIIth International Biometrics Conference, Hamilton, Ontario. pp. 19–33.
- Harville, D.A. (1997). Matrix Algebra from a Statisticians Perspective. First ed., Springer, New York, NY.

- Higdon, D., Swall, J., & Kern, J. (1999). Non stationary spatial modeling. First ed., Oxford University Press, Oxford.
- Hoef, J.V., & Barry, R. (1998). Modelling crossvariograms for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69, 275–294.
- Hoeting, J.A. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, 19, 574–577.
- Junttila, V., Kauranne, T., & Leppänen, V. (2010). Estimation of forest stand parameters from airborne laser scanning using calibrated plot databases. *Forest Science*, 56, 257–270.
- Kamper, T.U., Johnson, B.R., Kuester, M., & Keller, M. (2010). Neon: The first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing*, , 043510.
- Monnet, J.M., Chanussot, J., & Berger, F. (2011). Support vector regression for the estimation of forest stand parameters using airborne laser scanning. *IEEE Geoscience and Remote Sensing Letters*, 8, 580–584.
- Muss, J.D., Mladenoff, D.J., & Townsend, P.A. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, 115, 824 – 835.
- Niemann, K.O., Quinn, G., Goodenough, D.G., Visintini, F., & Loos, R. (2012). Addressing the effects of canopy structure on the remote sensing of foliar chemistry of a 3-dimensional, radiometrically porous surface. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5, 584–593.
- Popescu, S.C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment*, 115, 2786–2797.
- Reich, B.J., & Fuentes, M. (2007). A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1, 249–264.
- Salas, C., Ene, L., Gregoire, T.G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, 114, 1277 – 1285.
- Schmidt, A.M., & Gelfand, A.E. (2003). A bayesian coregionalization model for multivariate pollutant data. *Journal of Geophysics Research-Atmospheres*, 108, 8783.
- Sendak, P.E., Brissette, J.C., & Frank, R.M. (2003). Silviculture affects composition, growth, and yield in mixed northern conifers: 40-year results from the penobscot experimental forest. *Canadian Journal of Forest Research*, 33, 2116–2128.

Tonolli, S., Dalponte, M., Neteler, M., Rodeghiero, M., Vescovo, L., & Gianelle, D. (2011). Fusion of airborne lidar and satellite multispectral data for the estimation of timber volume in the southern alps. *Remote Sensing of Environment*, 115, 24862498.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Third ed., Springer Verlag, Berlin.

©2013 IEEE. Reprinted, with permission, from Chad Babcock, Jason Matney, Andrew O. Finley, Aaron Weiskittel and Bruce D. Cook, Multivariate Spatial Regression Models for Predicting Individual Tree Structure Variables Using LiDAR Data, *IEEE Journal of Selected Topics In Applied Earth Observations and Remote Sensing*, February 2013.

## CHAPTER 3

# IMPROVING LIDAR BASED PREDICTION OF FOREST BIOMASS USING HIERARCHICAL MODELS WITH SPATIALLY VARYING COEFFICIENTS

### 3.1 Introduction

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

Regression models proposed for forest AGB mapping using LiDAR data often do not explicitly accommodate residual spatial dependence, see, e.g., Anderson et al. (2008), Gonzalez et al. (2010), Muss et al. (2011), Tonolli et al. (2011), and Popescu et al. (2011). A non-spatial model can be appropriate if all spatially structured variation in the outcome is accounted for by the covariates used for model fitting; however, this is often an unrealistic assumption when data are spatially indexed. Considering forest attribute data, it is reasonable to expect similar outcomes for neighboring locations. For example, inventory plots close together could comprise like tree species and have comparable stem densities due to similar disturbance histories and environmental conditions. In contrast, inventory plots far apart are less likely to share common composition or structure attributions. A key assumption of regression is that model residuals are not correlated. Given the spatially dependent nature of forest inventory data, the potential for spatially correlated residuals in AGB models is high. Not accounting for residual spatial dependence can result in falsely precise estimates of model parameters and erroneous predictions (Hoeting, 2009). For these reasons, it is

important to check for spatially structured residuals when spatial data are used to fit AGB models.

To ensure a model’s statistical validity, spatial dependence among residuals can be accommodated by introducing a spatially varying intercept via the addition of an appropriately structured random effect. Including spatial random effects that account for spatial association between observed locations can not only help to reduce correlation among residuals but also improve model fit and prediction accuracy via partitioning error uncertainty to a spatial and non-spatial component and borrowing information from observed locations to inform prediction at proximate locations.

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

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

distributions of model parameters.

In an effort to more fully account for patterns of spatial dependence between AGB and LiDAR covariates, we assess the utility of a Bayesian hierarchical modeling framework that accommodates both residual spatial dependence and non-stationarity of model covariates through the introduction of spatial random effects.

We explore this objective using four forest inventory datasets that are part of the North American Carbon Program (NACP) each comprising point-referenced measures of AGB and discrete LiDAR. For each dataset, we consider a set of regression model specifications of varying complexity. Models are assessed based on fit criteria and predictive performance using a ten-fold cross-validation approach.

The remainder of the paper evolves as follows. The motivating datasets and field and LiDAR data processing are detailed in Section 3.2. The proposed modeling framework and model assessment are described in Section 3.3, followed by analysis results and associated discussion presented in Section 3.4. Finally, some concluding remarks with an indication of future work are provided in Section 3.5.

## 3.2 Study Sites

### 3.2.1 Fraser Experimental Forest

The Fraser Experimental Forest (FEF) is located in central Colorado (39° 4' N, 105° 52' W) near the town of Fraser. Tree species at FEF consist primarily of *Abies lasiocarpa* and *Picea engelmannii* at higher elevations and *Pinus contorta* at lower elevations. Climate at FEF is characterized by cold and relatively long winters, with mean annual temperature and precipitation of 0°C and 737 mm, respectively. FEF experienced a widespread stand-replacing fire in approximately 1685. Selective clearcuts were conducted as watershed-scale manipulations at FEF in the 1950s. Although FEF is currently experiencing mortality due

to mountain pine beetle, the field measurements and LiDAR acquisition for this study were completed prior to beetle infestation.

### 3.2.2 Marcell Experimental Forest

Located in northern Minnesota, the Marcell Experimental Forest (MEF) consists of mixed forests that include both upland forests and peatlands. Upland forests are generally dominated by *Populus tremuloides* and *grandidentata*, but contain substantial components of *Betula papyrifera*, *Pinus resinosa*, *Pinus strobus*, and *Pinus banksiana*. Lowland tree species include *Larix laricina*, *Picea mariana*, *Fraxinus nigra*, and *Thuja occidentalis*. Climate at MEF is subhumid continental, with mean annual precipitation of 785 mm, mean annual temperature of 3°C and air temperature extremes of  $-46^{\circ}\text{C}$  and  $38^{\circ}\text{C}$ . Forests of the Lake States region experienced widespread logging around the turn of the 20th century, including much of the MEF landscape, and natural disturbances at MEF include windstorms of variable intensity and rare wildfires.

### 3.2.3 Glacier Lake Ecosystem Experimental Site

Glacier Lake Ecosystem Experimental Site (GLEES) is located approximately 55 km west of Laramie, Wyoming at 3190 m elevation ( $41^{\circ} 22' \text{ N}$ ,  $106^{\circ} 14' \text{ W}$ ). GLEES includes a mosaic of trees and alpine meadows; the forest is dominated by *Picea engelmannii* and *Abies lasiocarpa*. GLEES has a mean annual temperature of  $-2^{\circ}\text{C}$  and a mean annual precipitation of 1000 mm, primarily as snow. Tree ages at GLEES suggest either a stand-replacing disturbance more than 400 years ago with slow recovery, or a series of smaller disturbances over the last 400 years (Bradford et al., 2008).

### 3.2.4 Niwot Long Term Ecological Research Site

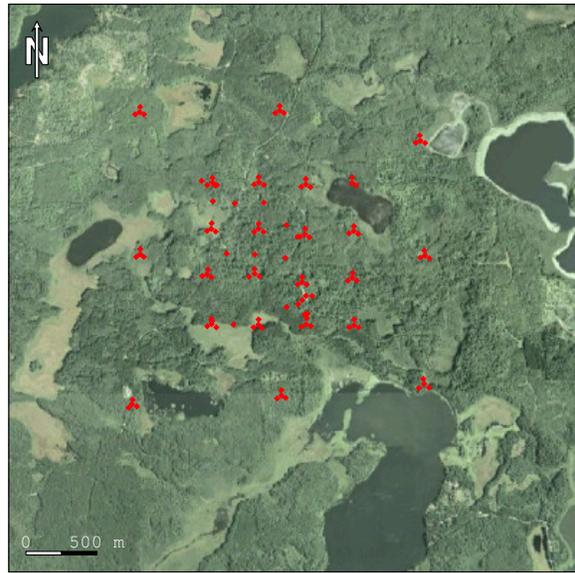
Niwot Long Term Ecological Research Site (NIWOT) is located at 3050 m elevation on the front range of the Rocky Mountains (40° 2' N, 105° 33' W), near the town of Nederland, Colorado. Tree species include primarily a mix of *Abies lasiocarpa*, *Picea engelmannii* and *Pinus contorta* with minor components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

### 3.2.5 Field Data Preparation

Field data at each site were collected using methods similar to the Forest Inventory and Analysis style plot design (Bechtold & Patterson, 2005). Each plot location consists of four subplots with radius 8–10 m (depending on site) with one being in the center and 3 others 35 m away from the plot center at 0°, 120°, and 240°. Additional plots consisting of only one subplot were also used in this analysis. FEF, MEF, NIWOT and GLEES contained 61, 115, 62 and 46 subplots respectively. Figure 3.1 shows the spatial distribution of the plots within the LiDAR coverage area. Within each subplot, individual tree diameter at breast height and height measurements for both live and dead trees were taken and used in species specific allometric equations to estimate AGB (stem, branch and foliage biomass). Additional details about field measurements and allometric equations used for biomass estimation are available in Bradford et al. (2010). The individual tree AGB estimates were totaled for each subplot and converted to AGB Mg/ha. The AGB measurements were then square root transformed to better approximate a normal distribution before model fitting.



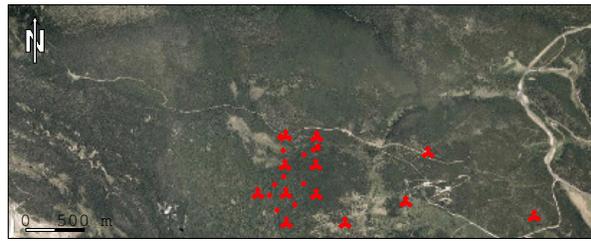
(a) FEF



(b) MEF



(c) GLEES



(d) NIWOT

Figure 3.1: Aerial photographs of FEF, MEF, GLEES, and NIWOT with inventory sub-plot locations highlighted in red.

### 3.2.6 LiDAR Data Preparation

Height-above-ground measurements for the first return pulses were calculated by subtracting the point elevations from a digital terrain model constructed from the LiDAR data. LiDAR return height empirical cumulative distribution curves were constructed over each sub-plot

and a percentile height dataset was compiled by extracting heights associated with 5% intervals (i.e., 5%–100%) for each curve. These percentile heights served as an initial set of regression covariates. Using the same intervals, a percentile height dataset was constructed for a  $20 \times 20$  meter grid over the LiDAR coverage area and subsequently used to construct AGB prediction maps. The cell size was chosen because it is approximately the same area as the observed sub-plots.

Variable selection via dimension reduction was necessary because of high collinearity among the percentile height covariates. We chose eigen (spectral) decomposition to accomplish this. Eigen decomposition is a matrix factorization technique that is useful for characterizing patterns in high-dimensional data (Harville, 1997). Eigen decomposition (or similar orthogonalization techniques) is used widely for data reduction or compression in statistics, signal processing, pattern recognition, remote sensing, and other fields where high-dimensional and/or highly correlated data are encountered (Guanter et al., 2012, 2013; Huang et al., 2014; Wang et al., 2014).

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

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

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

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

Let

$$\boldsymbol{\rho} = \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}'$$

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

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

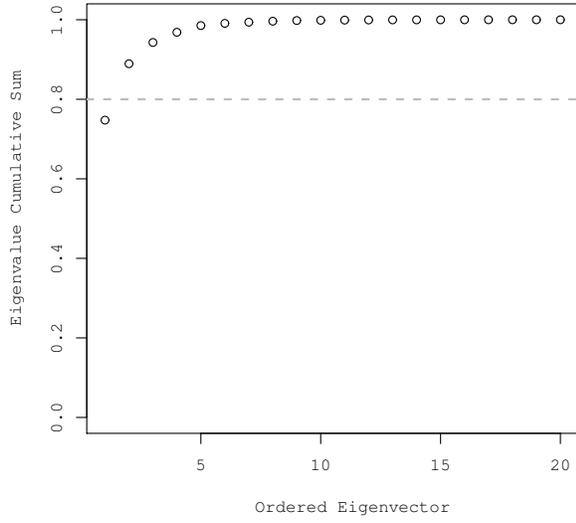
$$\tilde{\boldsymbol{B}}_{i,j} = \frac{\boldsymbol{B}_{i,j} - \boldsymbol{\mu}_j}{\boldsymbol{\sigma}_j}.$$

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

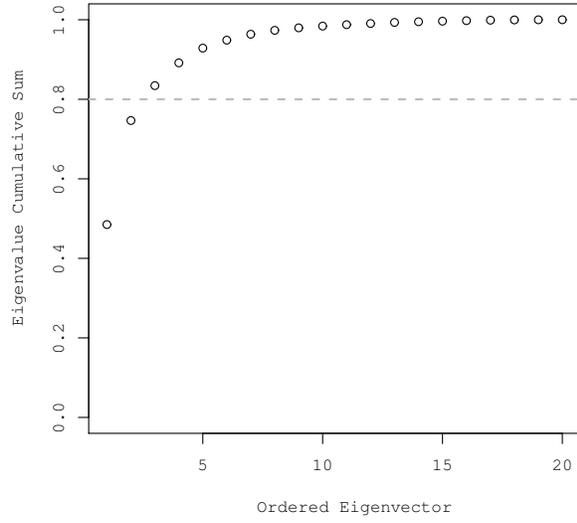
In practice, a subset of columns from  $\boldsymbol{A}^*$  (i.e., principal component scores) can be used as covariates in regression analysis—referred to as principal components regression (Chatterjee & Hadi, 2006). This is the approach we pursued here. The columns of  $\boldsymbol{A}^*$  represent a new, orthogonal (i.e., uncorrelated), candidate covariate set. From this set we retained the minimum number of covariates that explained at least 80% of the variance in the percentile height data. This was done by selecting the leftmost columns of  $\boldsymbol{A}^*$  whose standardized eigenvalues total was  $> 0.8$ . To meet this criterion the two leftmost columns of  $\boldsymbol{A}^*$  were retained for FEF, NIWOT and GLEES. It was necessary to keep three columns for MEF. Figure 3.2 shows a plot of the cumulative sum of the eigenvalues for each site. We kept the corresponding columns of  $\boldsymbol{B}^*$  for each site to use in subsequent grid prediction and mapping.

Some grid cell locations contained covariate values that well exceeded the range of values used for model fitting. This is a result of a relatively small set of observed field plots compared to the size and heterogeneity of the study area. To avoid potential issues arising from model extrapolation (Perrin, 1904), grid cells with values outside the support of the observed covariate values were removed from the analysis. In the AGB maps that follow,

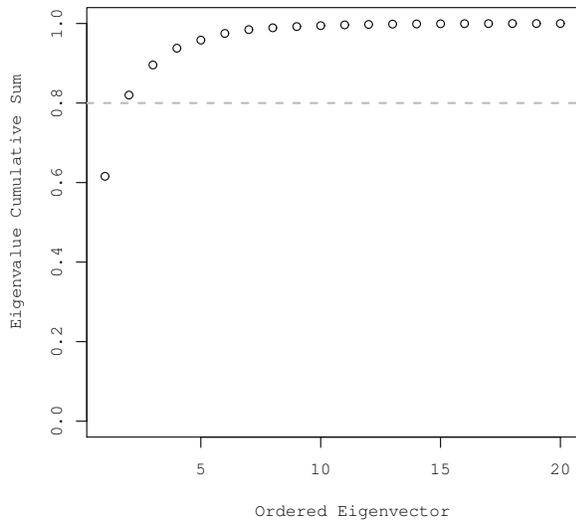
the excluded grid cells are shown in black.



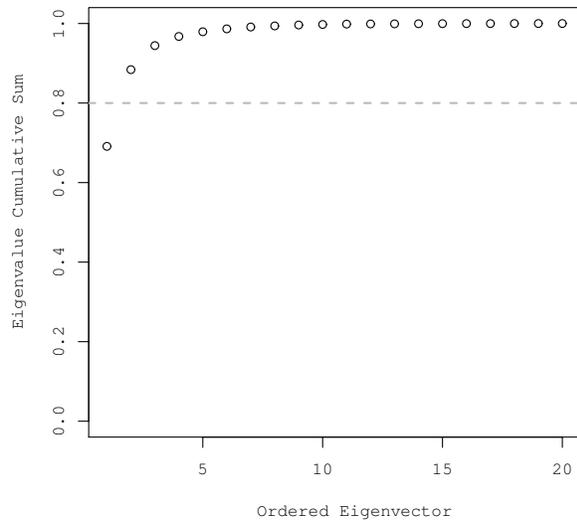
(a) FEF



(b) MEF



(c) GLEES



(d) NIWOT

Figure 3.2: The cumulative sums of the standardized eigenvalues (in decreasing order) for FEF, MEF, GLEES, and NIWOT.

### 3.3 Methods

#### 3.3.1 Modeling Framework

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

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s})'\mathbf{w}(\mathbf{s}) + \epsilon(\mathbf{s}). \quad (3.1)$$

Here the linear mean structure that accounts for large scale variation in the outcome is composed of a  $p \times 1$  vector  $\mathbf{x}(\mathbf{s})$ , comprising an intercept and spatially referenced covariates, and an associated column vector of regression coefficients  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ . The  $\tilde{\mathbf{x}}(\mathbf{s})$  is a  $q \times 1$  vector that includes the intercept and those covariates from  $\mathbf{x}(\mathbf{s})$  whose impact on the outcome is posited to vary spatially. This space varying impact is captured through the vector of spatial random effects  $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}), \dots, w_q(\mathbf{s}))'$ . Various sub-models are formed by specifying different combinations of  $\tilde{\mathbf{x}}(\mathbf{s})$  and associated  $\mathbf{w}(\mathbf{s})$ . Customarily,  $\epsilon(\mathbf{s})$  is modeled as an independent white-noise process that captures measurement error or micro-scale variation. With any collection of  $n$  locations, say  $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ , we assume that the  $\epsilon(\mathbf{s}_i)$ 's are identically and independently distributed as  $N(0, \tau^2)$ , where  $\tau^2$  is often called the *nugget*.

For this model, spatial structure is typically introduced through a multivariate Gaussian process (GP) (Cressie, 1993; Wackernagel, 2003; Banerjee et al., 2004), where a *covariance function* explicitly models the covariance of  $\mathbf{w}(\mathbf{s})$  both within and among locations. This additional flexibility is attractive in many settings (see, e.g., Gelfand et al., 2004; Banerjee et al., 2011; Finley et al., 2011); however, for the objectives of this study it is not clear the extra effort will fetch substantial advantages. Rather, for simplicity and substantially reduced computational demand, we assumed the elements of  $\mathbf{w}(\mathbf{s})$  arise from  $q$  independent univariate GPs. Specifically, the process associated with the  $k$ -th covariate

is  $w_k(\mathbf{s}) \sim GP(0, C(\cdot, \cdot; \boldsymbol{\theta}_k))$  where  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_k) = \text{Cov}(w_k(\mathbf{s}), w_k(\mathbf{s}^*))$  is a valid covariance function that models the covariance corresponding to a pair of locations  $\mathbf{s}$  and  $\mathbf{s}^*$ . The process realizations are collected into an  $n \times 1$  vector, say  $\mathbf{w}_k = (w_k(\mathbf{s}_1), \dots, w_k(\mathbf{s}_n))'$ , that follows a multivariate normal distribution  $MVN(\mathbf{0}, \Sigma_k)$ , where  $\Sigma_k$  is the  $n \times n$  covariance matrix of  $\mathbf{w}_k$  with  $(i, j)$ -th element given by  $C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_k)$ . Clearly  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_k)$  cannot be just any function; it must ensure that the resulting  $\Sigma_k$  matrix is symmetric and positive definite. Such functions are known as positive definite functions and are characterized as the characteristic function of a symmetric random variable. Further technical details about positive definite functions can be found in Cressie (1993), Chilès & Delfiner (1999), and Banerjee et al. (2004).

We specify  $C(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}_k) = \sigma_k^2 \rho(\mathbf{s}, \mathbf{s}^*; \phi_k)$  with  $\boldsymbol{\theta}_k = \{\sigma_k^2, \phi_k\}$ ,  $\rho(\cdot; \phi_k)$  is a valid spatial correlation function, where  $\phi_k$  quantifies the rate of correlation decay and  $\text{Var}(w_k) = \sigma_k^2$ . For the subsequent analyses we assumed an exponential correlation function,  $\rho(\|\mathbf{s} - \mathbf{s}^*\|; \phi_k) = \exp(-\phi_k \|\mathbf{s} - \mathbf{s}^*\|)$ , where  $\|\mathbf{s} - \mathbf{s}^*\|$  is the Euclidean distance between the locations  $\mathbf{s}$  and  $\mathbf{s}^*$ .

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

Using notations similar to Gelman et al. (2013), we can write the posterior distribution of the model parameters as  $p(\Omega | \mathbf{y})$ , where  $\Omega = \{\boldsymbol{\beta}, \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q, \sigma_1^2, \dots, \sigma_q^2, \phi_1, \dots, \phi_q, \tau^2\}$

and  $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))'$ , which is proportional to

$$\begin{aligned} \prod_{k=1}^q \text{Unif}(\phi_k | a_{\phi_k}, b_{\phi_k}) &\times \prod_{k=1}^q \text{IG}(\sigma_k^2 | a_{\sigma_k}, b_{\sigma_k}) \times N(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \Sigma_\beta) \times \text{IG}(\tau^2 | a_\tau, b_\tau) \times \\ &\times \prod_{k=1}^q N(\mathbf{w}_k | \mathbf{0}, \Sigma_k) \times \prod_{i=1}^n N(y(\mathbf{s}_i) | \mathbf{x}(\mathbf{s}_i)' \boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s})' \mathbf{w}(\mathbf{s}), \tau^2). \end{aligned} \quad (3.2)$$

An efficient MCMC algorithm for estimation of (3.2) is obtained by updating  $\boldsymbol{\beta}$  from its full conditional then using Metropolis steps for the remaining parameters. Alternatively, the model can be reparametrized such that the spatial random effects  $\mathbf{w}$  do not need to be sampled directly (Banerjee et al., 2008). In either case, the MCMC algorithms yield posterior samples of  $\Omega$ .

For predictions, if  $\mathcal{S}_0 = \{\mathbf{s}_{0,1}, \mathbf{s}_{0,2}, \dots, \mathbf{s}_{0,m}\}$  is a collection of  $m$  *new* locations, the posterior predictive distribution of the spatial random effects associated with the  $k$ -th regression coefficient is given by

$$p(\mathbf{w}_{k,0} | \mathbf{y}) \propto \int p(\mathbf{w}_{k,0} | \mathbf{w}_k, \Omega, \mathbf{y}) p(\mathbf{w}_k | \Omega, \mathbf{y}) p(\Omega | \mathbf{y}) d\Omega d\mathbf{w}_k, \quad (3.3)$$

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

Given  $L$  posterior samples,  $\{\Omega^{(l)}\}_{l=1}^L$ , this distribution can be obtained via composition sampling by first draw  $\mathbf{w}_k^{(l)}$  and then draw  $\mathbf{w}_{k,0}^{(l)}$  for each  $l$  from  $p(\mathbf{w}_{k,0} | \mathbf{w}_k^{(l)}, \Omega^{(l)}, \mathbf{y})$ , where this last distribution is derived as a conditional distribution from a multivariate normal and, hence, is multivariate normal. More precisely, the process realizations over the new locations are conditionally independent of the observed outcomes given the realizations over the observed locations and the process parameters. In other words,  $p(\mathbf{w}_{k,0} | \mathbf{w}_k, \Omega, \mathbf{y}) = p(\mathbf{w}_{k,0} | \mathbf{w}_k, \Omega)$ , which is a multivariate normal distribution with mean and variance given

by

$$\begin{aligned} E[\mathbf{w}_{k,0} | \mathbf{w}_k, \Omega] &= \text{Cov}(\mathbf{w}_{k,0}, \mathbf{w}_k) \text{Var}^{-1}(\mathbf{w}_k) \mathbf{w}_k = R_0(\phi_k)' R(\phi_k)^{-1} \mathbf{w}_k \\ \text{and } \text{Var}[\mathbf{w}_{k,0} | \mathbf{w}_k, \Omega] &= \sigma_k^2 \left\{ R(\phi_k) - R_0(\phi_k)' R(\phi_k)^{-1} R_0(\phi_k) \right\}, \end{aligned}$$

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

### 3.3.2 Candidate Models

Five candidate models were derived from (3.1) and include: *non-spatial*, where  $\mathbf{w}_k$ 's are set to zero; spatially varying intercept (*SVI*), where only the spatial random effects associated with the model intercept are included; the full spatially varying coefficients (*SVC*), where all regression coefficients have associated spatial random effects; *SVC- $\beta_1$* , where spatial random effects for the intercept and  $\beta_1$  are included, and; *SVC- $\beta_2$* , where spatial random effects for the intercept and  $\beta_2$  are included. Due to MEF's additional covariate, four additional candidate models were tested. *SVC- $\beta_3$*  has spatial random effects associated with  $\beta_0$  and  $\beta_3$ ; *SVC- $\beta_1\beta_2$*  has spatial random effects associated with  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ ; *SVC- $\beta_1\beta_3$*  has spatial random effects associated with  $\beta_0$ ,  $\beta_1$ , and  $\beta_3$ , and; *SVC- $\beta_2\beta_3$*  has spatial random effects associated with  $\beta_0$ ,  $\beta_2$ , and  $\beta_3$ .

Empirical semivariograms fit to the residuals of the *non-spatial* models (Figure 3.3) were used to help guide hyperprior specification for the candidate models' *IG* and *Unif* priors. Specifically, for the variance parameters we set the *IG* hyperprior  $a=2$ , which results in a prior distribution mean equal to  $b$  and infinite variance (see, e.g., *IG* definition in Gelman et al. (2013)). Then the  $b$  hyperpriors for the models'  $\tau^2$  and  $\sigma^2$ 's were set according to the

nugget and partial sill of the semivariograms. The prior for the spatial decay parameters,  $\phi$ 's, was set to  $Unif(0.0006, 3)$  which, assuming the exponential spatial correlation function, corresponds to support for an *effective spatial range* between  $\sim 1$ -5000 m. We define the effective spatial range as the distance at which the correlation equals 0.05.

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

Three MCMC chains were run for 25000 iterations each. The most demanding model required  $\sim 1$  hour to complete a single MCMC chain. Convergence was diagnosed using the CODA package in R by monitoring mixing of chains and the Gelman-Rubin statistic (Gelman & Rubin, 1992). Satisfactory convergence was diagnosed within 10000 iterations for all models. Posterior inference was based on a post burn-in sub-sample of 15000 iterations (5000 from each chain).

### 3.3.3 Fit and Prediction Accuracy Assessment

We assessed model performance using the popular Deviance Information Criterion ( $DIC$ ) to rank models in terms of how well they fit the data (Spiegelhalter et al., 2002). This criteria is the sum of the Bayesian deviance (a measure of model fit) and the effective number of parameters (a penalty for model complexity),  $D$  and  $p_D$ , respectively. Lower values of  $DIC$  indicate better model fit.

Predictive performance was assessed using a ten-fold cross-validation approach. Here, the full dataset was split into ten roughly equal sized subsets. Then AGB was predicted for locations within each subset, given parameters estimated from the remaining subsets. Root mean squared prediction error (RMSPE) was then calculated using the observed AGB values and corresponding median of the posterior predictive distribution. Lower RMSPE values signify more accurate predictions.

### 3.4 Results

Our experimental design produced 24 combinations of candidate models and sites. Tables 3.1, 3.2, 3.3, and 3.4 provide summaries of the candidate models' parameter estimates, fit, and predictive performance for FEF, MEF, GLEES and NIWOT, respectively. Concerning all sites *non-spatial* models, the exclusion of zero from the regression slope parameters' 95% credible interval suggests the derived LiDAR covariates explain a substantial portion of variability in AGB with the exception of NIWOT's  $\beta_2$  parameter.

The results in Tables 3.1, 3.3, and 3.4 show that for the FEF, GLEES, and NIWOT sites, the *SVI* models produce marginally to moderately lower  $D$ ,  $DIC$ , and  $RMSPE$  compared to the *non-spatial* models. This suggests the addition of a spatial random effect to the model intercept results in improved fit and predictive performance. Given the clear spatial structure seen in the exploratory semivariogram plots of the *non-spatial* models' residuals (Figure 3.3 (a), (c), and (d)) the SVIs' improved fit to the data is not surprising. These semivariogram plots along with the formal assessment of model fit and predictive performance suggest the *non-spatial* models violate the assumption of independent and identically distributed residuals at FEF, GLEES, and NIWOT.

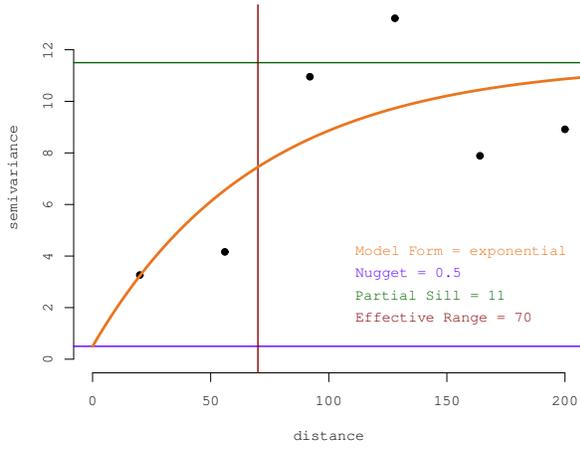
On the other hand, the semivariogram of the *non-spatial* model residuals for MEF (Figure 3.3 (b)), does not provide strong evidence of any residual spatial correlation. We also see in Table 3.2 only slight gains in model fit along with a decrease in model prediction accuracy when comparing the *non-spatial* and *SVI* candidates. This suggests the addition of a spatial random effect to the intercept at MEF is not only unnecessary, but allowing the intercept to vary spatially causes the model to over-fit the data and hence prediction suffers.

These observations are further corroborated by the candidate models' parameter estimates provided in Tables 3.1, 3.2, 3.3, and 3.4. The ratio  $\tau^2/(\tau^2 + \sigma_0^2)$  calculated using estimates from the *SVI* model is useful for assessing the need for spatial random effects.

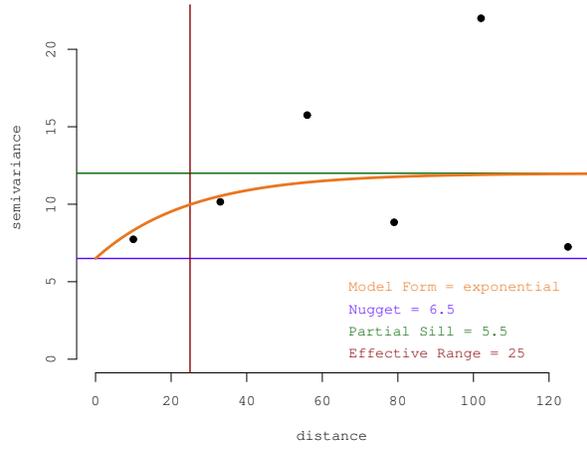
The lower the ratio, the greater the spatial structure in the *non-spatial* models' residuals and, hence, need to include the spatially-varying intercept. This ratio is 0.15, 0.13, and 0.29 for FEF, GLEES, and NIWOT, respectively. Compared to MEF's 0.45 ratio, we can see that the *non-spatial* model residuals for FEF, GLEES, and NIWOT contain more spatial structure.

The effective spatial range is another parameter that describes the strength of a spatial component. A long spatial range suggests stronger and farther reaching spatial dependence. We see from Tables 3.1, 3.2, 3.3, and 3.4 that MEF's *SVI* spatial range of 29.5 m is much shorter than the *SVI* ranges for FEF, GLEES and NIWOT. The long spatial ranges and low nugget to total variance ratios suggest the spatial random effects are capturing substantial residual spatial structure and the non-spatial models are not appropriate for the FEF, GLEES, and NIWOT datasets. In contrast, these same measures suggest there is little spatial structure in the MEF *non-spatial* model residuals and hence spatial random effects may not be warranted.

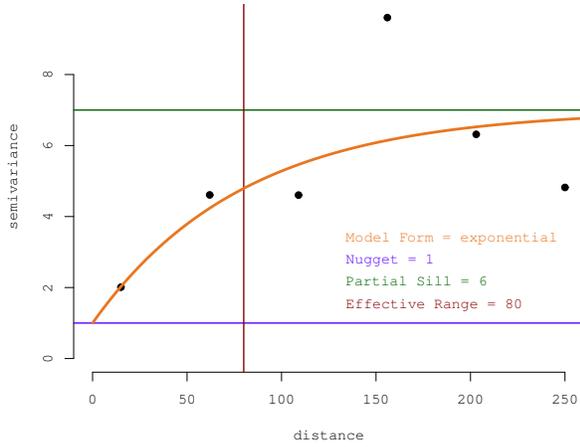
The remainder of this section is divided into four subsections, one for each site. In each subsection we identify which model was selected as the "best" and explore some site specific results.



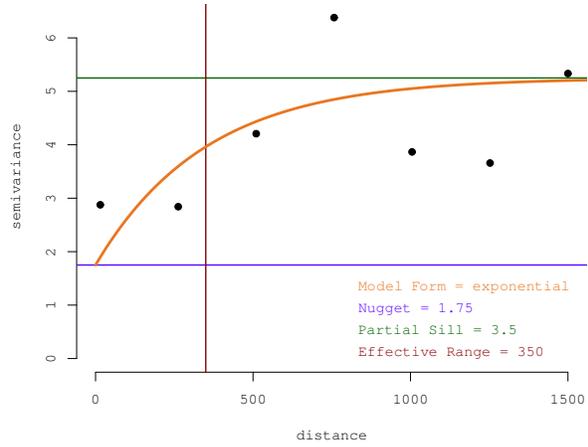
(a) FEF



(b) MEF

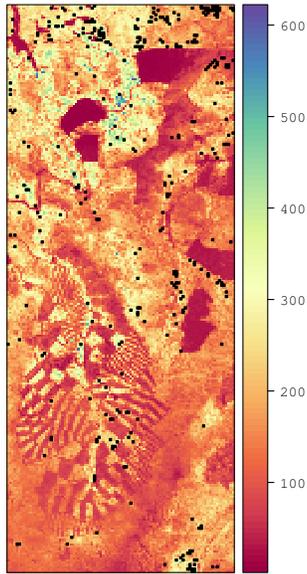


(c) GLEES

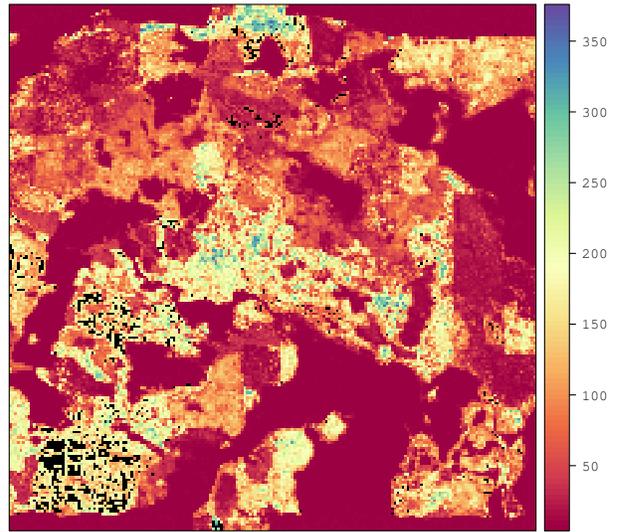


(d) NIWOT

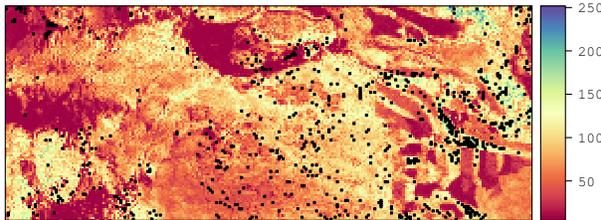
Figure 3.3: Semivariogram models generated using the residuals from the *non-spatial* model estimates for FEF, MEF, GLEES, and NIWOT.



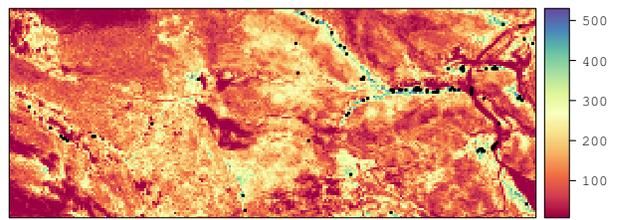
(a) FEF



(b) MEF

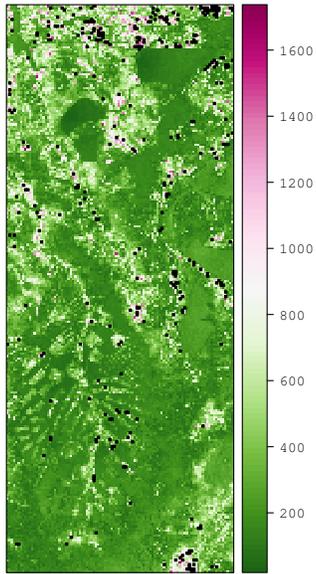


(c) GLEES

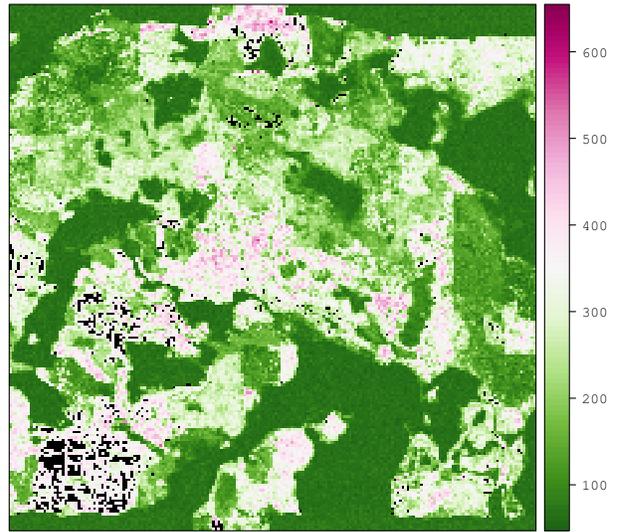


(d) NIWOT

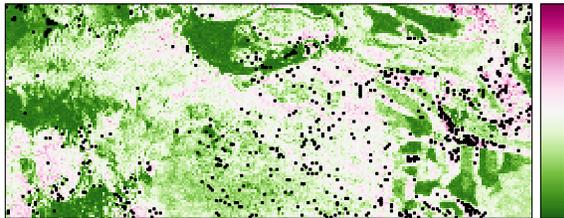
Figure 3.4: Maps showing predicted AGB means for the full LiDAR extent at FEF, MEF, GLEES, and NIWOT.



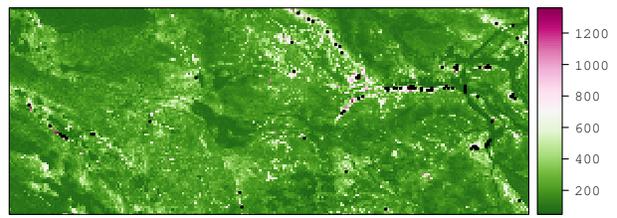
(a) FEF



(b) MEF



(c) GLEES



(d) NIWOT

Figure 3.5: Maps showing predicted AGB 95% credible interval width (a measure of prediction uncertainty) at FEF, MEF, GLEES, and NIWOT.

### 3.4.1 FEF Results

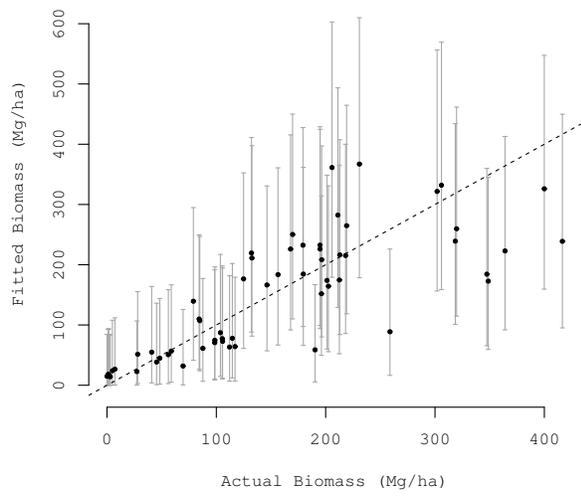
At FEF the full *SVC* model yields the lowest  $D$ ,  $DIC$ , and  $RMSPE$ . The *SVC* model shows a 83% improvement in  $DIC$  (model fit) compared to the *non-spatial* model. This improvement is clearly seen when comparing Figures 3.6a and 3.6b. Figure 3.6b shows high correspondence between the observed and fitted values for the *SVC* model, where as in Figure 3.6a, the fitted values are scattered further from the one-to-one line. It is also evident that the credible intervals for the *SVC* fitted values are much tighter than for the *non-spatial* fitted values. There is a 9% improvement in  $RMSPE$  (prediction accuracy) when moving from the *non-spatial* to *SVC* model. Figures 3.4a and 3.5a show maps of *SVC* model predicted biomass and prediction uncertainty, respectively. Most of the predictions in Figure 3.4a fall within the range of observed AGB (0 – 430 Mg/ha); however, those pixels that exceed this observed range exhibit high associated uncertainties (Figure 3.5a).

In addition to the regression coefficient estimates given in Table 3.1, it is also instructive to look at surfaces of  $\beta_k(\mathbf{s}) = \beta_k + w_k(\mathbf{s})$  which provide insight into the space-varying relationship between the  $k$ -th covariate and AGB. For example, the  $\beta_1(\mathbf{s})$  coefficient map (Figure 3.7b) shows clear spatial structure and suggests the associated LiDAR covariate explains more variability in AGB in the northern regions of the study area, i.e.,  $\beta_1(\mathbf{s})$  is further from zero in the north. Compared to  $\beta_1(\mathbf{s})$ , random effects associated with  $\beta_0(\mathbf{s})$  and  $\beta_2(\mathbf{s})$  have much shorter effect ranges (Table 3.1) which make their adjustments less noticeable at the scale shown in Figure 3.7. In regions far from inventory sub-plot locations the random effects on  $\beta_0$  and  $\beta_2$  retreat to zero. By zooming into a region with a high concentration of sub-plot inventories, switching from an equal to quantile interval color classification, and passing the predictions of the  $\beta(\mathbf{s})$ 's through a smoothing filter we can highlight the more subtle spatial adjustments to  $\beta_0$  and  $\beta_2$ . Figures 3.8a and 3.8b depict  $\beta_0(\mathbf{s})$  and  $\beta_2(\mathbf{s})$  at the extents shown by the red outline in Figures 3.7a and 3.7c, respectively. These surfaces show how the global estimates of  $\beta_0$  and  $\beta_2$ , given in Table 3.1 as  $\sim 11.28$  and  $\sim 0.91$  respec-

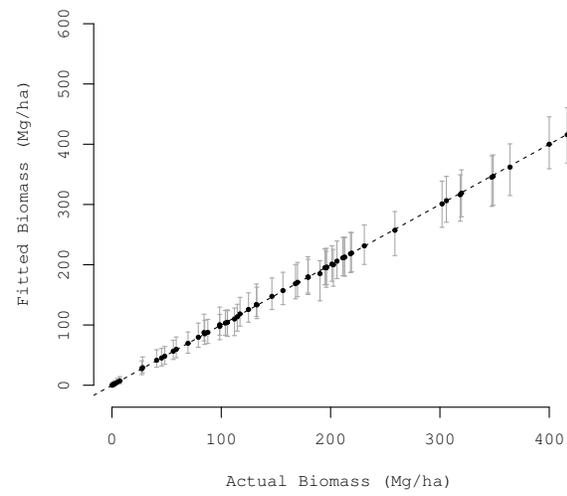
tively, are influenced by their corresponding spatial random effects to better accommodate the non-stationary relationship between the covariates and AGB. These adjustments, along with the increases in prediction accuracy and model fit support the use of the *SVC* model for FEF.

Table 3.1: Candidate model parameter estimates and 95% credible intervals for FEF.

		Non-Spatial		SVI		SVC- $\beta_1$		SVC- $\beta_2$		SVC	
parameter C.I. 50% (2.5%, 97.5%)	$\beta_0$	11.16	(10.49, 11.80)	11.33	(10.44, 12.27)	11.58	(10.88, 12.30)	11.41	(10.69, 12.05)	11.28	(10.69, 11.90)
	$\beta_1$	-1.08	(-1.25, -0.92)	-1.06	(-1.23, -0.88)	-1.04	(-1.91, 0.17)	-0.97	(-1.16, -0.78)	-0.81	(-1.76, 0.12)
	$\beta_2$	1.10	(0.70, 1.51)	1.01	(0.57, 1.44)	0.77	(0.35, 1.22)	1.46	(0.63, 3.51)	0.91	(0.35, 1.41)
	$\tau^2$	6.49	(4.62, 9.58)	0.89	(0.32, 3.70)	3.38	(1.61, 5.70)	2.36	(0.79, 4.83)	0.07	(0.02, 0.75)
	$3/\phi_0$	—		121.45	(20.05, 284.91)	21.14	(10.29, 241.36)	31.29	(10.37, 1612.90)	20.62	(10.27, 156.90)
	$3/\phi_1$	—		—		2380.95	(160.09, 4838.71)	—		2400.00	(302.72, 4918.03)
	$3/\phi_2$	—		—		—		194.74	(10.82, 2542.37)	21.63	(10.25, 185.87)
	$\sigma_0^2$	—		5.18	(2.51, 8.71)	0.48	(0.21, 1.73)	0.53	(0.21, 2.19)	0.48	(0.20, 1.19)
	$\sigma_1^2$	—		—		0.66	(0.29, 1.91)	—		0.74	(0.33, 1.95)
	$\sigma_2^2$	—		—		—		2.28	(0.80, 7.99)	1.64	(0.89, 3.09)
fit statistics	D	293.21		93.13		248.42		224.71		22.43	
	pD	3.93		38.24		24.17		29.57		25.71	
	<b>DIC</b>	<b>297.14</b>		<b>131.37</b>		<b>272.59</b>		<b>254.28</b>		<b>48.14</b>	
	<b>RMSPE</b>	<b>71.70</b>		<b>69.44</b>		<b>71.16</b>		<b>74.88</b>		<b>65.27</b>	

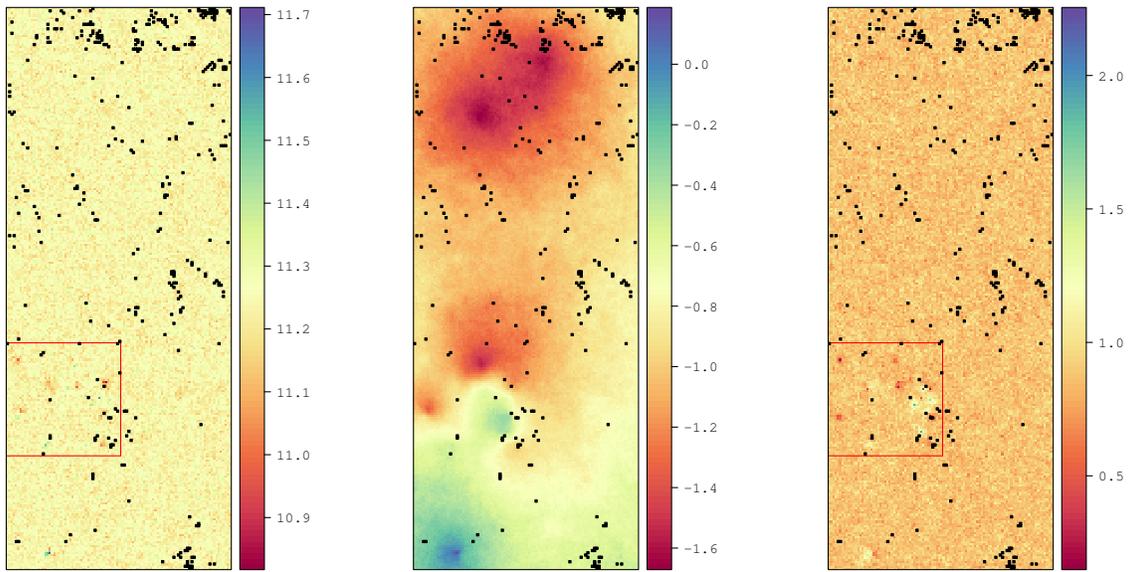


(a) *non-spatial*



(b) *SVC*

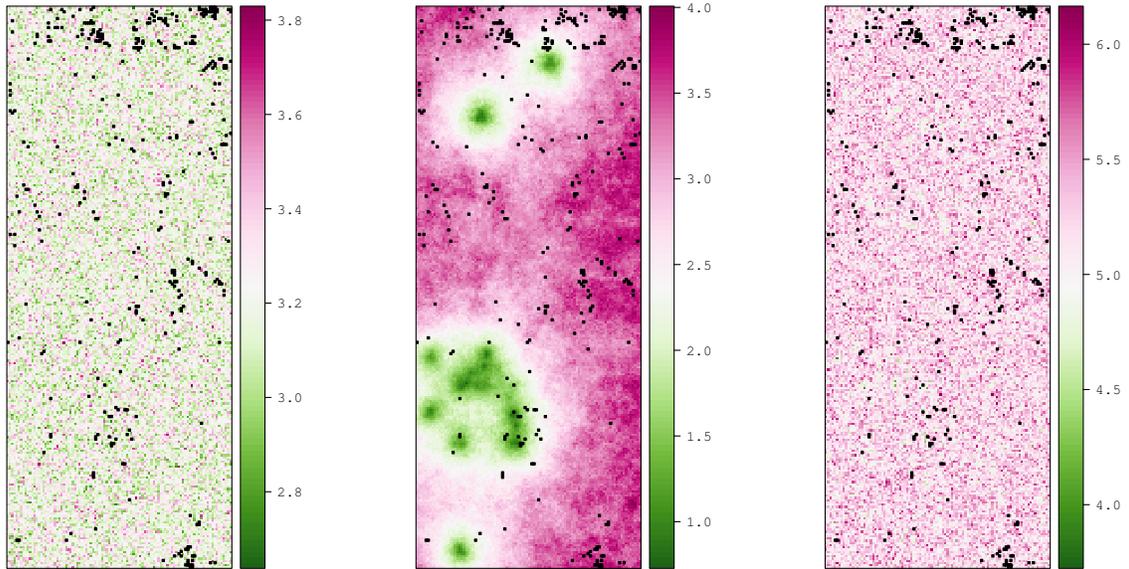
Figure 3.6: Plots showing fitted versus actual AGB values for FEF's *non-spatial* and *SVC* models. Vertical grey error bars depict 95% credible intervals for fitted values.



(a)  $\hat{\beta}_0 + \hat{w}_0$

(b)  $\hat{\beta}_1 + \hat{w}_1$

(c)  $\hat{\beta}_2 + \hat{w}_2$

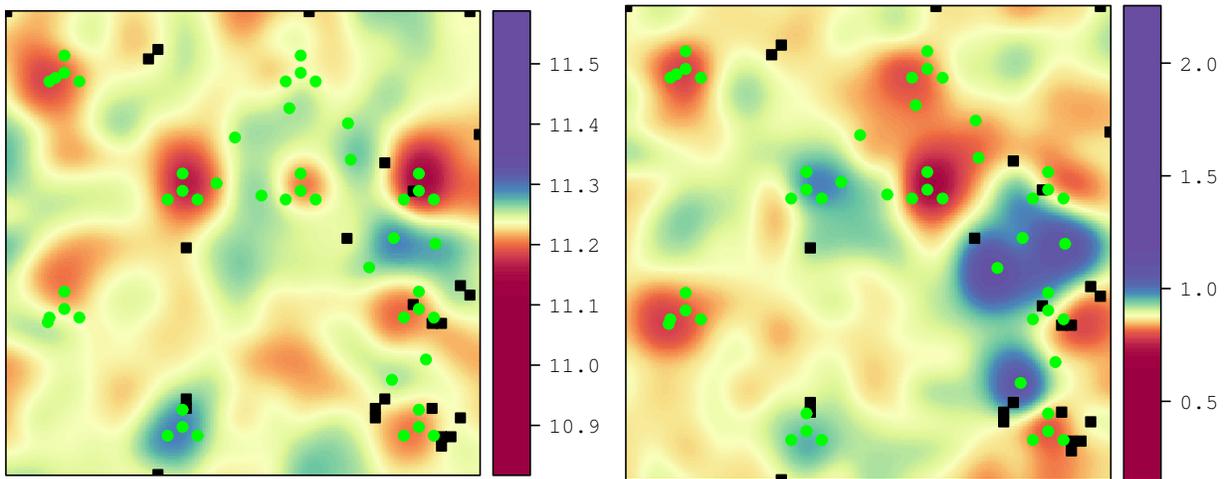


(d)  $\hat{\beta}_0 + \hat{w}_0$  95% CIW

(e)  $\hat{\beta}_1 + \hat{w}_1$  95% CIW

(f)  $\hat{\beta}_2 + \hat{w}_2$  95% CIW

Figure 3.7: Maps showing the spatially varying coefficients  $(\hat{\beta}_x + \hat{w}_x)$  and associated 95% credible interval width (CIW) for the *SVC* model at FEF.



(a)  $\hat{\beta}_0 + \hat{w}_0$

(b)  $\hat{\beta}_2 + \hat{w}_2$

Figure 3.8: Maps showing the spatially varying coefficients  $\hat{\beta}_0 + \hat{w}_0$  and  $\hat{\beta}_2 + \hat{w}_2$  zoomed into the extent outlined in red in figures 3.7a and 3.7c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green.

### 3.4.2 MEF Results

As noted previously, of the eight spatial models fit to the MEF dataset none out performed the *non-spatial* model with regard to prediction accuracy and only the *SVI* model shows an improvement in *DIC* over the *non-spatial* model (Table 3.2).

Figures 3.4b and 3.5b show the *non-spatial* model AGB prediction and associated uncertainty for the extent of the LiDAR data. A large number of grid cells in the southwest corner of MEF were removed from the prediction dataset because their covariate values far exceeded those seen in the observed dataset, see Section 3.2.6. We also see that grid cells in this area that were included in the prediction set have high uncertainties. The aerial photo in Figure 3.1b suggests that the forest in this region of MEF shows very different qualities than the areas where field data were collected.

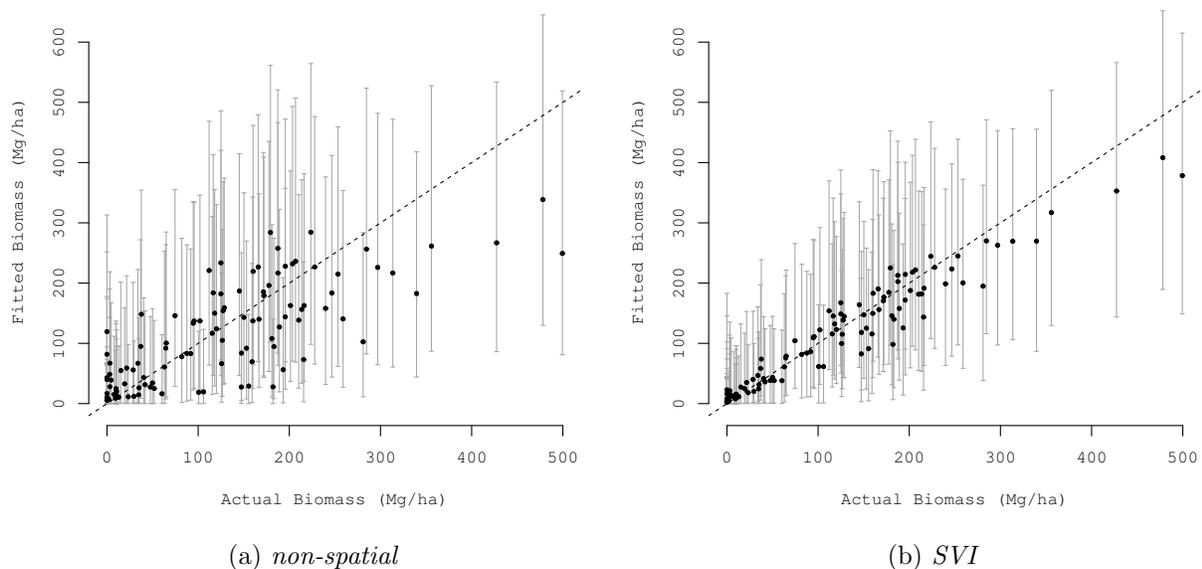


Figure 3.9: Plots showing fitted versus actual AGB values for MEF’s *non-spatial* and *SVI* models. Vertical grey error bars depict 95% credible intervals for fitted values.

Table 3.2: Candidate model parameter estimates and 95% credible intervals for MEF.

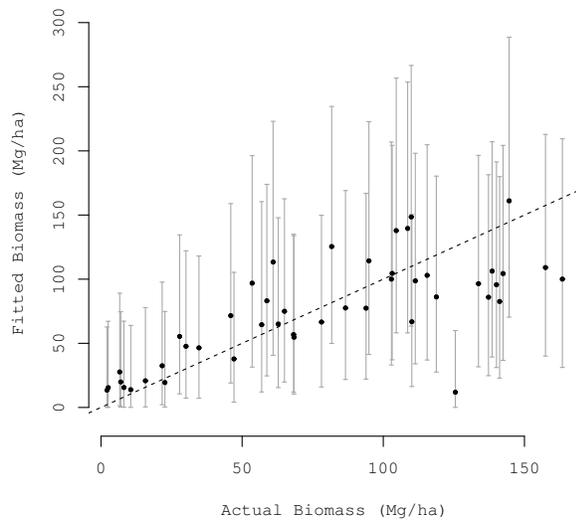
	Non-Spatial	Spatial	SVC- $\beta_1$	SVC- $\beta_2$	SVC- $\beta_3$	SVC- $\beta_1\beta_2$	SVC- $\beta_1\beta_3$	SVC- $\beta_2\beta_3$	SVC	
parameter C.I. 50% (2.5%, 97.5%)	$\beta_0$	9.10 (8.45, 9.73)	9.08 (8.43, 9.73)	9.00 (8.30, 9.69)	9.12 (8.44, 9.80)	9.09 (8.45, 9.73)	9.01 (8.27, 9.75)	8.98 (8.27, 9.69)	9.09 (8.39, 9.79)	9.00 (8.22, 9.75)
	$\beta_1$	-1.29 (-1.49, -1.09)	-1.29 (-1.48, -1.09)	-1.23 (-1.50, -0.93)	-1.28 (-1.51, -1.06)	-1.28 (-1.49, -1.06)	-1.22 (-1.51, -0.91)	-1.22 (-1.54, -0.84)	-1.27 (-1.51, -1.03)	-1.21 (-1.52, -0.91)
	$\beta_2$	0.89 (0.61, 1.15)	0.88 (0.61, 1.15)	0.90 (0.57, 1.23)	0.83 (0.35, 1.29)	0.86 (0.53, 1.18)	0.84 (0.30, 1.38)	0.88 (0.48, 1.25)	0.82 (0.33, 1.30)	0.83 (0.27, 1.38)
	$\beta_3$	-1.20 (-1.67, -0.74)	-1.18 (-1.64, -0.71)	-1.26 (-1.84, -0.70)	-1.25 (-1.83, -0.68)	-1.21 (-1.85, -0.59)	-1.34 (-2.03, -0.64)	-1.29 (-2.00, -0.60)	-1.29 (-2.00, -0.57)	-1.38 (-2.18, -0.60)
	$\tau^2$	11.56 (8.99, 15.36)	4.98 (1.24, 11.25)	10.51 (7.75, 14.37)	10.15 (7.55, 13.75)	10.30 (7.67, 13.93)	9.52 (6.67, 13.44)	9.27 (6.92, 13.59)	9.85 (7.16, 13.58)	9.10 (6.19, 13.07)
	$3/\phi_0$	—	29.50 (12.04, 192.98)	20.33 (10.26, 284.90)	21.51 (10.28, 607.29)	22.44 (10.29, 928.79)	20.30 (10.29, 332.59)	21.23 (10.29, 417.25)	21.02 (10.29, 572.52)	20.96 (10.27, 466.56)
	$3/\phi_1$	—	—	24.52 (10.32, 1612.90)	—	—	25.30 (10.32, 1363.64)	32.31 (10.36, 2439.02)	—	25.08 (10.31, 594.06)
	$3/\phi_2$	—	—	—	22.33 (10.31, 699.30)	—	21.35 (10.29, 518.13)	—	22.06 (10.32, 584.80)	22.29 (10.32, 390.12)
	$3/\phi_3$	—	—	—	—	19.52 (10.25, 276.50)	—	20.07 (10.25, 323.62)	19.59 (10.26, 386.25)	19.76 (10.27, 250.63)
	$\sigma_0^2$	—	5.98 (1.31, 11.37)	0.36 (0.17, 1.02)	0.36 (0.17, 0.90)	0.36 (0.17, 1.01)	0.36 (0.17, 1.01)	0.36 (0.17, 0.96)	0.37 (0.17, 1.05)	0.36 (0.17, 0.92)
	$\sigma_1^2$	—	—	0.36 (0.19, 0.69)	—	—	0.39 (0.21, 0.80)	0.39 (0.20, 0.80)	—	0.40 (0.21, 0.79)
	$\sigma_2^2$	—	—	—	0.65 (0.30, 1.52)	—	0.70 (0.32, 1.71)	—	0.64 (0.28, 1.52)	0.70 (0.31, 1.78)
	$\sigma_3^2$	—	—	—	—	0.99 (0.37, 2.75)	—	1.08 (0.41, 3.23)	0.94 (0.38, 2.77)	1.07 (0.41, 3.10)
fit statistics	D	608.41	501.53	600.15	595.76	597.53	588.21	590.97	592.34	583.05
	pD	4.95	41.29	28.61	22.85	19.05	41.54	37.34	29.88	47.55
	DIC	<b>613.37</b>	<b>542.82</b>	<b>628.77</b>	<b>618.61</b>	<b>616.59</b>	<b>629.74</b>	<b>628.31</b>	<b>622.22</b>	<b>630.60</b>
	RMSPE	<b>66.89</b>	<b>67.46</b>	<b>67.53</b>	<b>70.81</b>	<b>68.23</b>	<b>72.41</b>	<b>68.44</b>	<b>73.59</b>	<b>75.28</b>

### 3.4.3 GLEES Results

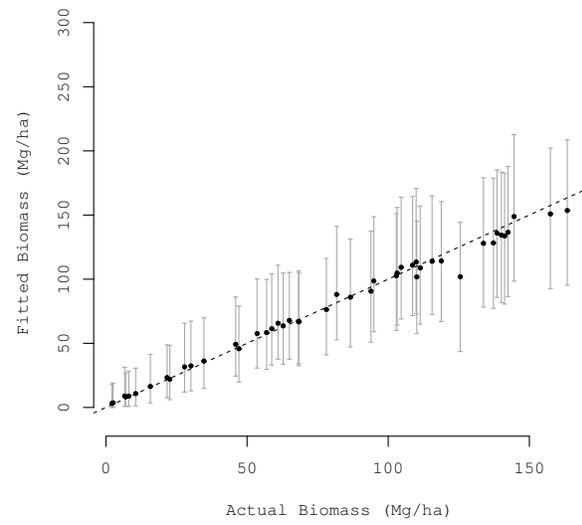
The *SVI* model at GLEES has the lowest *RMSPE* value of the five models (Table 3.3); however, we see lower *D* and *DIC* values in all three *SVC* variant models. Again the increased flexibility afforded by the additional spatial random effects in the *SVC* models improves the fit but may cause prediction accuracy to suffer due to over-fitting. For this reason the *SVI* model is preferred for the GLEES dataset. We see a 33% improvement in *DIC* and a 4.5% increase in prediction accuracy moving from the *non-spatial* to the *SVI* model. Figure 3.10b shows stronger one-to-one correspondence between fitted and observed AGB values comparing to the models non-spatial counterpart in Figure 3.10a.

Table 3.3: Candidate model parameter estimates and 95% credible intervals for GLEES.

		Non-Spatial		SVI		SVC- $\beta_1$		SVC- $\beta_2$		SVC	
parameter C.I. 50% (2.5%, 97.5%)	$\beta_0$	8.27	(7.66, 8.86)	8.40	(7.76, 9.14)	8.31	(7.61, 9.03)	7.94	(7.43, 8.51)	8.06	(7.38, 8.74)
	$\beta_1$	-0.62	(-0.79, -0.44)	-0.63	(-0.81, -0.45)	-0.61	(-0.89, -0.31)	-0.61	(-0.77, -0.46)	-0.62	(-0.96, -0.34)
	$\beta_2$	0.66	(0.37, 0.96)	0.67	(0.37, 0.99)	0.86	(0.49, 1.23)	0.74	(0.14, 1.29)	0.87	(0.33, 1.45)
	$\tau^2$	4.10	(2.75, 6.63)	0.53	(0.19, 2.01)	0.40	(0.10, 1.67)	0.32	(0.10, 1.22)	0.25	(0.08, 0.95)
	$3/\phi_0$	—		44.11	(12.90, 173.16)	22.85	(10.31, 125.63)	20.96	(10.27, 151.98)	20.19	(10.27, 198.41)
	$3/\phi_1$	—		—		33.88	(10.46, 527.24)	—		29.80	(10.37, 1020.50)
	$3/\phi_2$	—		—		—		58.09	(10.56, 552.49)	28.21	(10.39, 643.78)
	$\sigma_0^2$	—		3.45	(1.95, 5.88)	0.98	(0.35, 2.66)	0.71	(0.29, 1.77)	0.53	(0.24, 1.33)
	$\sigma_1^2$	—		—		0.44	(0.24, 0.88)	—		0.39	(0.21, 0.85)
	$\sigma_2^2$	—		—		—		1.11	(0.54, 2.48)	0.78	(0.35, 1.90)
fit statistics	D	196.30		103.69		90.35		79.59		67.79	
	pD	3.90		29.79		27.58		30.25		33.74	
	<b>DIC</b>	<b>200.20</b>		<b>133.48</b>		<b>117.93</b>		<b>109.85</b>		<b>101.53</b>	
	<b>RMSPE</b>	<b>35.79</b>		<b>34.21</b>		<b>35.63</b>		<b>35.02</b>		<b>35.20</b>	



(a) *non-spatial*



(b) *SVI*

Figure 3.10: Plots showing fitted versus actual AGB values for GLEES's *non-spatial* and *SVI* models. Vertical grey error bars depict 95% credible intervals for fitted values.

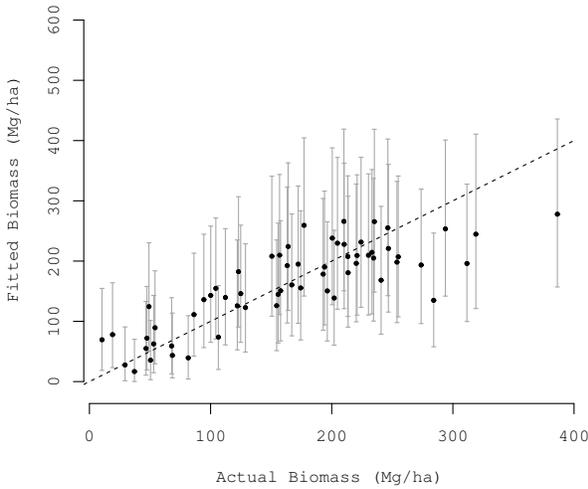
#### 3.4.4 NIWOT Results

The  $SVC-\beta_2$  model provides the best fit and prediction for the NIWOT dataset. This model yields a 24% lower  $DIC$  and 7% lower  $RMSP E$  value compared to the *non-spatial* model. As with GLEES, the more complex  $SVC$  model provides better fit, but the increased flexibility results in decreased predictive power. Comparing Figure 3.11a and 3.11b reveals greater one-to-one correspondence and tighter credible intervals for the  $SVC-\beta_2$  over that of the *non-spatial* model.

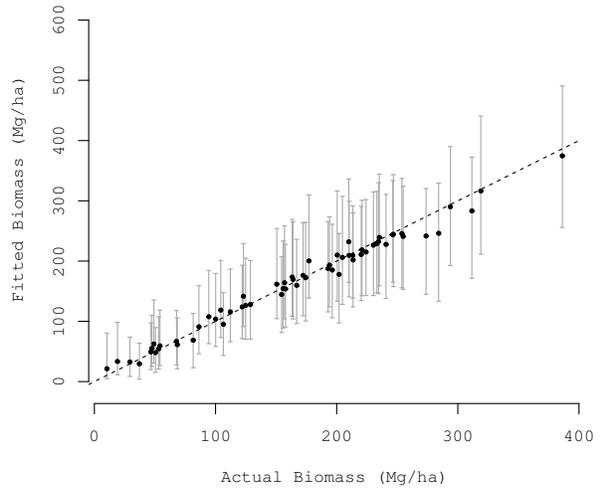
Predicted biomass and associated uncertainty maps for NIWOT are provided in Figures 3.4d and 3.5d, respectively. As in GLEES and FEF, AGB values predicted outside the range of observed AGB have the largest associated uncertainties. Figures 3.12a and 3.12c show the space varying impact of  $w_0$  and  $w_2$  on  $\beta_0$  and  $\beta_2$ , respectively. Zooming into the region of observed data and applying a similar smoothing filter as used for the FEF surfaces, we can again see local parameter adjustment to accommodate the non-stationary impact of the covariates on AGB (Figure 3.13). As mentioned previously,  $\beta_2$  for NIWOTs' *non-spatial* model is insignificant. When  $\beta_2$  is allowed to vary across the domain, as in  $SVC-\beta_2$ , we see that even though globally insignificant, in local regions the covariate can explain AGB variability and contribute to an overall increase in prediction accuracy.

Table 3.4: Candidate model parameter estimates and 95% credible intervals for NIWOT.

	Non-Spatial	SVI	SVC- $\beta_1$	SVC- $\beta_2$	SVC	
parameter C.I. 50% (2.5%, 97.5%)	$\beta_0$	12.34 (11.83, 12.85)	12.35 (11.38, 13.18)	12.35 (11.42, 13.01)	12.19 (11.38, 12.85)	12.21 (11.47, 13.06)
	$\beta_1$	-0.82 (-0.95, -0.68)	-0.77 (-0.93, -0.60)	-0.84 (-1.17, -0.17)	-0.77 (-0.93, -0.59)	-0.84 (-1.22, -0.29)
	$\beta_2$	0.21 (-0.04, 0.49)	0.21 (-0.05, 0.48)	0.31 (-0.01, 0.62)	0.23 (-0.52, 0.78)	0.34 (-0.17, 0.82)
	$\tau^2$	4.01 (2.82, 5.98)	1.29 (0.36, 3.74)	1.09 (0.25, 2.76)	0.99 (0.21, 3.14)	0.66 (0.20, 2.10)
	$3/\phi_0$	—	116.58 (13.64, 1444.64)	60.19 (10.52, 1578.95)	137.24 (11.40, 1167.32)	42.98 (10.43, 1401.87)
	$3/\phi_1$	—	—	50.58 (10.40, 2097.90)	—	77.34 (10.46, 2419.35)
	$3/\phi_2$	—	—	—	45.68 (10.45, 2343.75)	22.29 (10.30, 669.64)
	$\sigma_0^2$	—	2.66 (0.78, 5.07)	1.19 (0.34, 3.21)	1.74 (0.41, 3.87)	0.87 (0.30, 2.51)
	$\sigma_1^2$	—	—	0.37 (0.20, 1.03)	—	0.38 (0.20, 1.00)
	$\sigma_2^2$	—	—	—	0.74 (0.32, 2.03)	0.78 (0.33, 1.90)
fit statistics	D	262.40	189.65	177.34	172.77	150.60
	pD	3.93	27.65	34.78	28.97	41.52
	<b>DIC</b>	<b>266.32</b>	<b>217.29</b>	<b>212.12</b>	<b>201.74</b>	<b>192.12</b>
	<b>RMSPE</b>	<b>50.05</b>	<b>48.27</b>	<b>49.31</b>	<b>46.31</b>	<b>49.40</b>

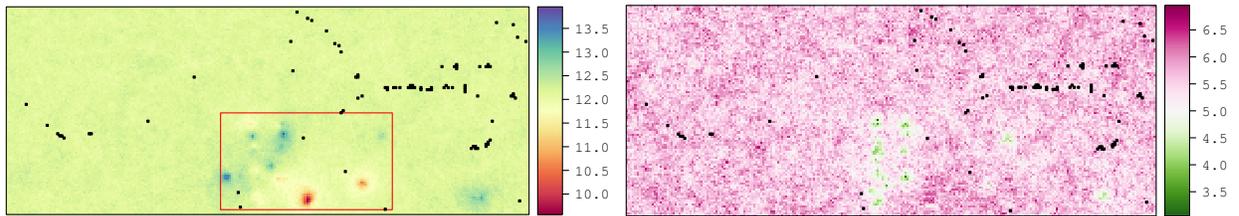


(a) *non-spatial*



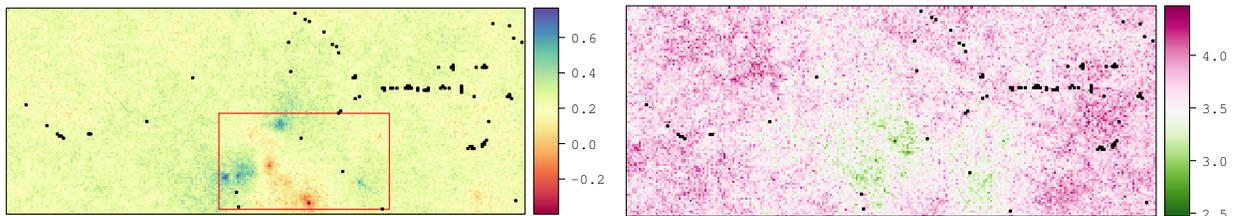
(b) *SVC- $\beta_2$*

Figure 3.11: Plots showing fitted versus actual AGB values for NIWOT's *non-spatial* and *SVC- $\beta_2$*  models. Vertical grey error bars depict 95% credible intervals for fitted values.



(a)  $\hat{\beta}_0 + \hat{w}_0$

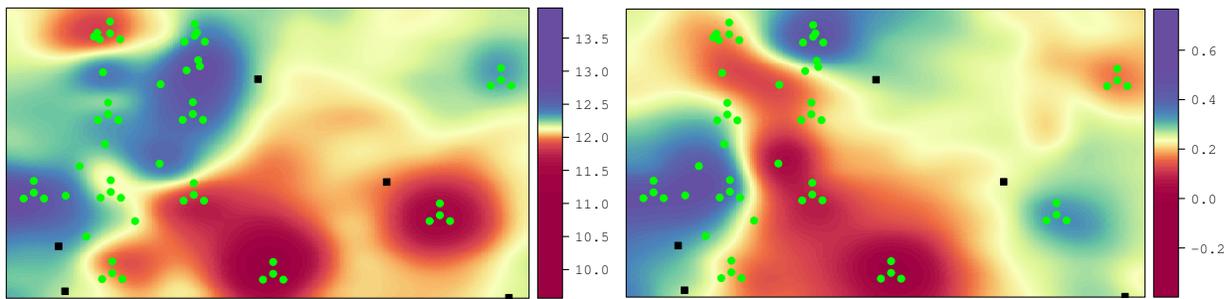
(b)  $\hat{\beta}_0 + \hat{w}_0$  95% CIW



(c)  $\hat{\beta}_2 + \hat{w}_2$

(d)  $\hat{\beta}_2 + \hat{w}_2$  95% CIW

Figure 3.12: Maps showing the spatially varying coefficients ( $\hat{\beta}_x + \hat{w}_x$ ) and associated 95% credible interval width (CIW) for the *SVC*- $\beta_2$  model at NIWOT.



(a)  $\hat{\beta}_0 + \hat{w}_0$

(b)  $\hat{\beta}_2 + \hat{w}_2$

Figure 3.13: Maps showing the spatially varying coefficients  $\hat{\beta}_0 + \hat{w}_0$  and  $\hat{\beta}_2 + \hat{w}_2$  zoomed into the extent outlined in red in figures 3.12a and 3.12c. A quantile interval color classification and smoothing filter are used to highlight the more subtle spatial adjustments of these parameters. The sub-plot locations are identified in green.

### 3.5 Discussion

For three of the four study sites, the *non-spatial* model residuals exhibit enough spatial correlation to warrant the use of a spatial random effect on the intercept. The presence of these serial correlations among the residuals suggest the assumptions of the non-spatial model are violated and the validity of subsequent inference is questionable. Further, this suggests that beyond meeting model assumptions, explicitly accommodating the residual dependence via spatially structured random effects might fetch improved prediction by borrowing information from proximate observed locations. This was illustrated in the analysis of the FEF, GLEES, and NIWOT datasets, where the *SVI* models outperformed the *non-spatial* models.

Uncertainty quantification is an important component of carbon Measurement, Reporting, and Verification systems (MRVs) like those called for by the United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries and actively being developed by NASA’s Carbon Monitoring System Science Team (UN-REDD, 2009; CMS, 2010). The uncertainty maps like the ones developed in Figure 3.5 can aid uncertainty analysis by providing a tool to assess the reliability of AGB predictions.

Mapping random effects can shed light on missing covariates. As an example, the random effect associated with  $\beta_1$  at FEF (Figure 3.7b) is moderately correlated with elevation. Although not explored here, after fitting a model as we did for FEF, a researcher could now consider elevation as an additional covariate either as an additive component to the model or test the multiplicative interaction with the LiDAR covariates.

Given the small size of the datasets considered here, the proposed modeling framework was computationally feasible. However, when the datasets consist of several thousand observations, which is common in forest inventory databases, cubic order matrix operations required for evaluating the model likelihood make parameter estimation computationally

onerous. Therefore, our future work will focus on exploring strategies for dimension reduction when fitting spatially varying coefficients models. Further, we will extend the spatially varying coefficients models to accommodate multiple forest variables of interest.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Anderson, J., Plourde, L., Martin, M., Braswell, B., Smith, M.L., Dubayah, R., Hofton, M., & Blair, J. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112, 1856 – 1870.
- Banerjee, S., Carlin, C., & Gelfand, A. (2004). Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Chapman & Hall/CRC.
- Banerjee, S., Finley, A., Waldmann, P., & Ericsson, T. (2011). Hierarchical spatial process models for multiple traits in large genetic trials. *Journal of the American Statistical Association*, 105, 506 – 521.
- Banerjee, S., Gelfand, A., Finley, A., & Huiyan, S. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70, 825 – 848.
- Bechtold, W., & Patterson, P. (2005). The enhanced forest inventory and analysis program—national sampling design and estimation procedures. General technical report SRS, USDA Forest Service, Southern Research Station.
- Bradford, J., Birdsey, R., Joyce, L., & Ryan, M. (2008). Tree age, disturbance history, and carbon stocks and fluxes in subalpine rocky mountain forests. *Global Change Biology*, 14, 2882 – 2897.
- Bradford, J., Weishampel, P., Smith, M.L., Kolka, R., Birdsey, R., Ollinger, S., & Ryan, M. (2010). Carbon pools and fluxes in small temperate forest landscapes: Variability and implications for sampling design. *Forest Ecology and Management*, 259, 1245 – 1254.
- Carlin, B., & Louis, T. (2000). Bayes and Empirical Bayes Methods for Data Analysis, Second Edition. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.
- Chatterjee, S., & Hadi, A. (2006). Regression Analysis by Example, 4th Edition. Wiley Series in Probability and Statistics, John Wiley & Sons.
- Chilès, J., & Delfiner, P. (1999). Geostatistics: modeling spatial uncertainty. Wiley series in probability and statistics, John Wiley & Sons.
- CMS (2010). Nasa carbon monitoring system. <http://carbon.nasa.gov>. Accessed: 3-18-2014.
- Cressie, N. (1993). Statistics for spatial data. Wiley series in probability and mathematical statistics: Applied probability and statistics, John Wiley & Sons.

- Finley, A. (2010). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2, 143 – 154.
- Finley, A., Banerjee, S., & MacFarlane, D. (2011). A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. *Journal of the American Statistical Association*, 106, 31 – 48.
- Finley, A., Banerjee, S., & McRoberts, R. (2009). Hierarchical spatial models for predicting tree species assemblages across large domains. *Annals of Applied Statistics*, 3, 1052 – 1079.
- Gelfand, A., Schmidt, A., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13, 263–312.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2013). Bayesian Data Analysis, 3rd Edition. Chapman & Hall/CRC Texts in Statistical Science, Chapman & Hall/CRC.
- Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- Gonzalez, P., Asner, G., Battles, J., Lefsky, M., Waring, K., & Palace, M. (2010). Forest carbon densities and uncertainties from lidar, quickbird, and field measurements in california. *Remote Sensing of Environment*, 114, 1561–1575.
- Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P., Gómez-Dans, J., Kuze, A., Suto, H., & Grainger, R. (2012). Retrieval and global assessment of terrestrial chlorophyll fluorescence from gosat space measurements. *Remote Sensing of Environment*, 121, 236–251.
- Guanter, L., Rossini, M., Colombo, R., Meroni, M., Frankenberg, C., Lee, J.E., & Joiner, J. (2013). Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space. *Remote Sensing of Environment*, 133, 52–61.
- Harville, D. (1997). Matrix Algebra from a Statistician’s Perspective. Springer.
- Hoeting, J. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, 19, 574–577.
- Huang, B., Song, H., Cui, H., Peng, J., & Xu, Z. (2014). Spatial and spectral image fusion using space matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 1693–1704.
- Lim, K., & Treitz, P. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research*, 19, 558–570.

- Lucas, R., Cronin, N., Lee, A., Moghaddam, A., Witte, C., & Tickle, P. (2006). Empirical relationships between airsar backscatter and lidar-derived forest biomass, queensland, australia. *Remote Sensing of Environment*, 100, 407–425.
- Muss, J., Mladenoff, D., & Townsend, P. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, 115, 824–835.
- Perrin, E. (1904). On some dangers of extrapolation. *Biometrika*, 3, 99–103.
- Popescu, S., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment*, 115, 2786–2797.
- Sherrill, K., Lefsky, M., Bradford, J., & Ryan, M. (2008). Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central rockies. *Canadian Journal of Forest Research*, 38, 2081–2096.
- Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64, 583–639.
- Tonolli, S., Dalponte, M., Neteler, M., Rodeghiero, M., Vescovo, L., & Gianelle, D. (2011). Fusion of airborne lidar and satellite multispectral data for the estimation of timber volume in the southern alps. *Remote Sensing of Environment*, 115, 2486–2498.
- UN-REDD (2009). The un-redd programme. <http://www.un-redd.org/>. Accessed: 3-18-2014.
- Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*, 3rd edition. Springer.
- Wang, C., Yu, W., Wang, R., Deng, Y., & Zhao, F. (2014). Comparison of nonnegative eigenvalue decompositions with and without reflection symmetry assumptions. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 2278–2287.
- Wheeler, D., & Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, 9, 145–166.
- Wheeler, D., & Waller, L. (2009). Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, 11, 1–22.

## CHAPTER 4

### SUMMARY AND FUTURE WORK

This thesis explored the use of flexible Bayesian hierarchical regression frameworks for calibrating LiDAR remote sensing information to forest inventory datasets. Chapters 2 and 3 provide examples where using Bayesian hierarchical spatial models can improve fit and predictive performance when modeling both univariate and multivariate forest structure data. When model residuals exhibit spatial correlation, that relationship can be absorbed into a spatially structured random effect to ensure statistically valid model inference and increase the predictive ability of the model. Further, Chapter 2 shows how leveraging not only spatial relationships, but correlations among different variables within the same tree can improve predictive performance. Chapter 3 demonstrates how building models that accommodate coefficient non-stationarity via the introduction of spatial random effects can yield increased model fit and prediction accuracies when using LiDAR to model and map forest characteristics.

As the forestry and remote sensing research communities strive to answer questions about more complex and dynamic ecological associations among differing environmental variables across space and time, richer classes of models beyond ordinary least squares or maximum likelihood methods will be needed. Bayesian hierarchical models for both spatial and spatio-temporal processes are seeing rapid development in the applied environmental sciences because of their immense flexibility to suit intricate variable associations and the ease of propagation of uncertainty through complex hierarchies.

In future work, I plan to explore the use of Bayesian hierarchical modeling for forestry and ecological processes in more detail. The research in this thesis has shown the potential for spatial random effects to improve prediction accuracies. A next logical step is to examine

how spatial random effects can be used to uncover missing covariates. Exploring what latent processes the spatial random effects might be capturing can lead to more ecologically meaningful results by providing insight into the underlying environmental processes causing the spatial distribution of the outcome variable, or variables, of interest.

The study sites examined in this thesis cover small spatial extents, i.e., stand level. Particularly with the models explored in Chapter 3, spatially varying coefficients may play a more significant role for fitting models over larger domains possibly crossing climatic zones or having large altitude gradients. Changes in these types of macro environmental conditions over one spatial domain usually indicate the need to fit more than one model, i.e., divide the original extent into smaller regions and fit uniquely parametrized regressions to each. Allowing coefficients to vary over space allows for parameter influence to change moving from one region to another. Rather than having several models with potentially low numbers of sample locations, one model can be formed that is able to borrow strength across the observations of the full domain and be less likely to suffer inadequate sample size. Also, instead of imposing divisions of the original domain based on speculation, the space varying coefficients model allows the data to inform when and how parameters change.

As we move into a data rich era in the environmental sciences, there is increasing need for computationally efficient methods for estimating cumbersome models with rich parameter association structures. In the future, I plan to further develop and refine the computational procedures used in this thesis. I hope to eventually author a package for the statistical programming language R so that other researchers interested in ecological modeling can easily execute the statistical methods explored in Chapters 2 and 3 in their research. The study sites examined here were all reasonably sized which meant that the estimation of these models was computationally feasible. As environmental datasets grow larger, new, more computationally efficient fitting procedures are needed. In future work, I plan to look into speeding up computation time for large datasets by examining the use of increased computational power

given by new technologies, such as processor hyper-threading and parallel processing to make complex ecological modeling more accessible to the environmental research field.