Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2008 *Paper* 48

Space-Time Regression Modeling of Tree Growth Using the Skew-t Distribution

Farouk S. Nathoo*

*University of Victoria, nathoo@math.uvic.ca

http://biostats.bepress.com/cobra/art48

Space-Time Regression Modeling of Tree Growth Using the Skew-t Distribution

Farouk S. Nathoo

Abstract

In this article we present new statistical methodology for the analysis of repeated measures of spatially correlated growth data. Our motivating application, a ten year study of height growth in a plantation of even-aged white spruce, presents several challenges for statistical analysis. Here, the growth measurements arise from an asymmetric distribution, with heavy tails, and thus standard longitudinal regression models based on a Gaussian error structure are not appropriate. We seek more flexibility for modeling both skewness and fat tails, and achieve this within the class of skew-elliptical distributions. Within this framework, robust space-time regression models are formulated using random effect growth curves, with coefficients arising from an underlying multivariate spatial process. Computational difficulties arise when data are collected at a large number of locations, and we consider two approaches for spatial modeling in the large data context. Both approaches are compared within the context of our application, and inference is conducted in a Bayesian framework, with implementation based on hybrid Monte Carlo.

1. Introduction

The growth of an individual tree is a complex process, influenced by several interacting factors. Understanding the nature of tree growth is of great importance in forestry, for breeding program development, as well as in directing selections for seed orchards. To this end, longitudinal studies of forest ecology often collect data on tree growth over time. Here, repeated measures are taken across an even-aged stand, and the height of each tree is recorded over a sequence of time points. Interest will often lie in uncovering relationships between tree growth and other ecological processes, such as forest insect infestation, or in describing the variability associated with tree-specific characteristics, such as the genetic origin of trees.

Regression models for repeated measures growth data, particularly methods based on random effect growth curves (Zimmerman and Núñez-Antón, 2001; Diggle et al. 2002), are useful for the analysis of such longitudinal data. Typically, the subjects generating the longitudinal data are assumed independent, and a Gaussian assumption is made for analysis. In the forest ecology context, interaction through local properties of the landscape, arising from spatial variation in soil, topographic, geologic and micro-meteorological factors create spatial dependence in tree growth measurements. Moreover, this spatial structure is expected to evolve over time as a forest plantation ages, so that dynamic, spatially explicit models are required. In addition, data on height growth are often markedly non-Gaussian, exhibiting asymmetry and excess kurtosis, so that robust models accommodating these features are needed. Transformations to normality are possible in certain situations; however, this approach has been criticized, in particular for correlated data (see for example, Azzalini and Capitanio, 1999). Direct parametric modeling of the skewness and kurtosis through generalizations of the normal distribution is preferable, as this improves the interpretability of the regression model inferences.

In this article we present new statistical methodology for the analysis of repeated measures of spatially correlated growth data. We focus on parametric families capable of accommodating, and quantifying, departures from normality. Our motivating application, a ten year study of height growth in a plantation of juvenile white spruce, presents several challenges for statistical analysis. The longitudinal growth profiles arising from this study are depicted in Figure 1, panel (a), and the spatial locations of the trees generating the data are depicted panel (b). Here, the growth measurements arise from an asymmetric distribution, with heavy tails, and thus standard longitudinal regression models incorporating Gaussian errors are not strictly appropriate. We seek more flexibility for modeling both skewness and fat tails, and achieve this within the class of skew-elliptical distributions (Azzalini and Capitanio, 2003). Within this framework, space-time regression models are formulated using random effect growth curves, with coefficients arising from an underlying multivariate spatial

process. Computational difficulties arise when data are collected at a large number of locations, and we consider two approaches for spatial modeling in the large data context. The first, achieves dimension reduction through a discrete process convolution (Higdon, 1998), while the second is based on a Markov random field representation of the residual spatial structure (Besag et al., 1991). Both approaches are compared within the context of our application, with inference conducted in a Bayesian framework, and implementation based on hybrid Monte Carlo.

Modeling spatially correlated growth data using spatially-varying growth curves was recently considered by Banerjee and Johnson (2006), who develop weed growth models based on a linear specification. Spatial variability in the intercept and slope coefficients is modeled through a flexible bivariate spatial process, allowing for non-stationarity and multi-resolution spatial dependence. Importantly, the methods developed there are applicable under a Gaussian assumption for the observed data; whereas, we seek a more flexible model, as the Gaussian distribution is a poor representation of the mechanism generating growth data in many applications, such as the one considered here.

Flexible parametric modeling with skew-elliptical distributions has received considerable attention in the recent literature (see Genton, 2004). A general formulation of multivariate skew-elliptical densities has been described by Azzalini and Capitanio (2003), who develop maximum likelihood inference for the special case of regression models based on the multivariate skew-t distribution. Regression modeling with the univariate skew-t distribution has also been considered by Fernández and Steel (1998), under a slightly different formulation, with inference conducted in a Bayesian framework. For longitudinal data, linear mixed models based on skew-elliptical distributions have recently been proposed by Jara et al. (2008), who introduce skew-elliptical distributions for modeling both data and random effects, as well as by Ma et al. (2004), where a generalized skew-elliptical distribution is incorporated for modeling random effects at the second stage of a linear mixed model. For spatial data, there has been relatively little work in this area, an exception being Kim and Mallick (2004) who developed methods for spatial prediction based on a skew-Gaussian spatial process, focussing on asymmetric distributions. The methods developed there are not directly applicable to the analysis considered here. First, we must accommodate spatial data exhibiting not only skewness but also excess kurtosis. Second, our hierarchical framework incorporates multivariate spatial random effects corresponding to data observed over $n = 4330$ locations; a sample size which seems intractable under the formulation of Kim and Mallick (2004), based on the required large matrix computations.

We consider two approaches for modeling spatially correlated random effects in a computationally tractable manner. The first is based on a multivariate spatial process derived through process convolution (Higdon, 1998; Higdon, 2002; Calder, 2007; Zhou and Sansó, 2007). Process convolution models, also known as *spatial moving averages*, provide an intu-

itive construction for a spatial process. Importantly, the construction is useful in the analysis of large spatial data when used in conjunction with dimension-reduced approximations. These approximations have been studied by Kern (2000), and more recently by Xia and Gelfand (2005), who develop important results regarding stationary process approximation, based on a kernel solving approach. Our second formulation for spatial modeling is based on a Markov random field (Besag et al., 1991). Specifically, we assume that random effects arise under a multivariate generalization of the Gaussian intrinsic autoregression (Gamerman et al., 2002; Carlin and Banerjee, 2003). Here, the spatial distribution is constructed through a set of compatible full conditional distributions. This conditional specification ameliorates the computational difficulty associated with computing the inverse of a large covariance matrix, as the elements of this inverse matrix are modeled explicitly through the conditional specifications.

We specify our hierarchical models and conduct inference within a Bayesian framework, implemented through Markov chain Monte carlo (MCMC). Motivated by poor performance of MCMC samplers based on random walk updating, we develop an algorithm based on hybrid Monte Carlo (Gustafson, 1997). The hybrid algorithm works well in accelerating MCMC convergence with highly parameterized models, particularly when random walk exploration of the parameter space fails, or proves grossly inefficient.

The remainder of this paper is structured as follows. In Section 2 we review the univariate skew-normal and skew-t distributions. We discuss some properties of the skew-t distribution, which make it well suited for modeling data on tree growth, as well as its representation as a Gaussian mixture, which we shall exploit for model fitting. In Section 3, a formal hierarchical Bayes space-time modeling framework is proposed, and implementation based on hybrid Monte Carlo is discussed. Section 4 presents an analysis of the white spruce growth data. Beginning with an exploratory analysis, we demonstrate the strong need for non-Gaussian distributions and spatial modeling. Our methodology is then applied to the data, and the results obtained from fitting several hierarchical models are presented and compared. Section 5 concludes with a brief summary and discusses possible refinements to the current model.

2. Review of the Skew-normal and Skew-t Distributions

In this section we briefly review and summarize some basic concepts related to the skew-normal and skew-t distributions. Both distributions fall within the broad class of skew-elliptical distributions, and we refer the reader to Azzalini and Capitanio (2003) and Genton (2004) for an extensive treatment. A univariate random variable Y is said to have a skew-

normal distribution $Y \sim SN(\mu, \sigma^2, \alpha)$ with location $\mu \in \mathbf{R}$, scale $\sigma > 0$ and shape $\alpha \in \mathbf{R}$ if it is continuous, with density function

$$p(y|\mu, \sigma, \alpha) = \frac{2}{\sigma}\phi(\frac{y-\mu}{\sigma})\Phi(\frac{\alpha(y-\mu)}{\sigma}), \quad y \in \mathbf{R} \tag{1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the density and distribution function of the standard normal distribution. The shape parameter α governs the asymmetry of the distribution, with positive and negative values corresponding to positive and negative skewness respectively, and with $\alpha = 0$ corresponding to the $N(\mu, \sigma^2)$ distribution. Generalizing further, the standard construction of the t-distribution through the ratio of a normal random variable and the square-root of an independent χ^2 random variable divided by its degrees-of-freedom, can be extended to define a skew t-distribution via $Y = \mu + \frac{X}{\sqrt{W/\nu}}$ where $X \sim SN(0, \sigma^2, \alpha)$ and W is independently distributed as χ^2_ν, $\nu > 0$. The resulting four parameter distribution, denoted as $St(\mu, \sigma^2, \alpha, \nu)$, contains the normal ($\alpha = 0,\ \nu \to \infty$), skew-normal ($\nu \to \infty$) and t-distribution ($\alpha = 0$) as special cases, and has density given by

$$p(y|\mu, \sigma, \alpha, \nu) = 2t(y; \mu, \sigma, \nu)T\{\frac{\alpha(y-\mu)}{\sigma}(\frac{\nu+1}{\frac{(y-\mu)^2}{\sigma^2}+\nu})^{1/2}; \nu+1\} \tag{2}$$

where $t(y; \mu, \sigma, \nu) = \frac{\Gamma\{(\nu+1)/2\}}{\sigma(\pi\nu)^{1/2}\Gamma(\nu/2)}[1 + \frac{(y-\mu)^2}{\nu\sigma^2}]^{-(\nu+1)/2}$, the density of a t-distribution having location μ, scale σ and ν degrees of freedom; and $T(\cdot; \nu+1)$ is the cumulative distribution function of a standard t-distribution on $\nu+1$ degrees of freedom. The moments of the skew-t distribution (2) were derived by Azzalini and Capitanio (2003), and in particular, the mean and variance, when these exist, are given by

$$E[Y|\mu, \sigma, \alpha, \nu] = \mu + \frac{\sigma\alpha}{\sqrt{1+\alpha^2}}(\frac{\nu}{\pi})^{1/2}\frac{\Gamma\{\frac{1}{2}(\nu-1)\}}{\Gamma(\frac{1}{2}\nu)}, \quad \nu > 1 \tag{3}$$

$$Var[Y|\mu, \sigma, \alpha, \nu] = \sigma^2(\frac{\nu}{\nu-2} - \frac{\alpha^2}{1+\alpha^2}\frac{\nu}{\pi}\frac{\Gamma^2\{\frac{1}{2}(\nu-1)\}}{\Gamma^2(\frac{1}{2}\nu)}), \quad \nu > 2.$$

Finally, we note that the skew-t distribution (2) can be derived as a Gaussian mixture model, based on a three-stage hierarchical specification,

$$Y|Z, W \sim N(\mu + \frac{\alpha\sigma Z}{\sqrt{1+\alpha^2}}, \frac{\sigma^2}{W(1+\alpha^2)}) \tag{4}$$

$$Z|W \sim TN_{(0,\infty)}(0, W^{-1}), \quad W \sim \text{Gamma}(\nu/2, \nu/2)$$

$$\to Y \sim St(\mu, \sigma^2, \alpha, \nu)$$

where $TN_{(0,\infty)}(0, W^{-1})$ denotes a $N(0, W^{-1})$ distribution, truncated to the interval $(0, \infty)$. We shall make use of this representation for model fitting within a latent variable framework.

The skew-t distribution provides a flexible and intuitive departure from the normal distribution, allowing both asymmetry and heavy tails. We have found it necessary to accommodate these features, in addition to spatial correlation, in developing adequate regression models for longitudinal tree growth data.

3. Hierarchical Space-time Models

We develop our models within the context of our motivating application, which involves a study of tree growth in a plantation of white spruce. A primary question relates to assessing genetic variability in height growth. To this end, open-pollinated progenies from 142 families (maternal trees) were planted on the site in a randomized complete block design of $n_B = 8$ replicated blocks with four-tree row plots randomly assigned within each block. Tree mortality occurred in some families; thus, only $n_F = 139$ families with data from a total of $n = 4330$ trees were retained for analysis. Data for each tree having spatial location $\boldsymbol{s_i} \in \mathbf{D} \subset \mathbf{R^2}$ (depicted in Figure 1) are collected over a 10 year period. Specifically, for each tree, we obtain repeated height measurements over time, and we let $H_i(t_{ij})$ denote the height of the i^{th} tree at time t_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, m_i$. In addition, associated with each such measurement is a vector of explanatory variables $\boldsymbol{x}_i(t_{ij})$, and we let $F_i \in \{1, \ldots, n_F\}$ denote the family origin of the i^{th} tree, and $B_i \in \{1, \ldots, n_B\}$ denote the block in which the i^{th} tree was planted.

We specify a longitudinal regression model for height growth within a hierarchical Bayes framework, where, at the first level of the model, we assume that the heights are drawn from a skew-t distribution $H_i(t_{ij})|\mu_i(t_{ij}), \sigma, \alpha, \nu \overset{ind}{\sim} St(\mu_i(t_{ij}), \sigma^2, \alpha, \nu)$ where variability in height is modeled through the location parameter $\mu_i(t_{ij})$, and this variability is introduced through a mixed regression framework $\mu_i(t) = \boldsymbol{\beta'} \boldsymbol{x}_i(t) + \mathbf{b'} \boldsymbol{z}_\mathbf{i}(\mathbf{t})$. Here, $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ are vectors of covariates corresponding to fixed effects $\boldsymbol{\beta}$ and random effects $\boldsymbol{b}$ respectively. Specific to our application, we adopt a regression specification incorporating block and family random effects, in addition to a spatially-varying growth curve

$$\mu_i(t) = \boldsymbol{\beta'} \boldsymbol{x}_i(t) + b_{F_i}^{(F)} + b_{B_i}^{(B)} + \mu_{b0} + \mu_{b1} t + b_0(\boldsymbol{s_i}) + b_1(\boldsymbol{s_i}) t, \tag{5}$$

where $b_l^{(F)} \overset{iid}{\sim} N(0, \sigma_F^2)$, $l = 1, \ldots, n_F$, is a family random effect, with σ_F^2 characterizing genetic variation; $b_l^{(B)} \overset{iid}{\sim} N(0, \sigma_B^2)$, $l = 1, \ldots, n_F$, is a random block effect, with σ_B^2 characterizing the large scale variability across the blocks within the plantation; and $\mathbf{b}(\boldsymbol{s}) =$

$(b_0(\boldsymbol{s}), b_1(\boldsymbol{s}))'$ is a bivariate, zero-mean, Gaussian spatial process. The tree specific random effects $\mathbf{b}(\boldsymbol{s_i})$ operate over top of the fixed linear trend $\mu_{b0} + \mu_{b1}t$, allowing spatial variability in the growth trajectories, as well as inducing dependence in the longitudinal observations collected at each tree. In addition, a random effects growth curve allows for a heteroscedastic variance structure (when the variance exists) over time. That is, marginally, on integrating over the random effects, the model incorporates a time-dependent variance for the height measurements, and a with a linear growth curve, this variance is a quadratic function of time.

In specifying a model for the spatial random effects $(b_0(\boldsymbol{s_i}), b_1(\boldsymbol{s_i}))$, $i = 1, \ldots, n$, computational considerations arising from the large number $n = 4330$ of spatial locations become important. Direct specification of a bivariate Gaussian spatial process through a mean and a valid cross-covariance function is possible; however, this approach will lead to considerable difficulty for model fitting. We thus consider more constructive approaches in specifying a model for the spatial random effects. The first approach, following Higdon (2002), reduces the computational difficulty through dimension reduction, using a discrete process convolution. The second approach is based on a bivariate Markov random field. Specifically, we employ a bivariate intrinsic autoregression (Gamerman, 2003; Carlin and Banerjee, 2003). In the next two subsections we describe the two spatial modeling approaches in more detail.

3.1. Spatial Modeling with Discrete Process Convolutions

A stationary Gaussian spatial process specified through a process convolution (PC) is based on convolving a white noise process with an appropriately defined smoothing kernel. In this context, the spatially-varying intercept defined in (5) is based on the construction

$$b_0(\boldsymbol{s}) = \int_{\mathbf{R^2}} K_0(\boldsymbol{u} - \boldsymbol{s}) dX_0(\boldsymbol{u}) \tag{6}$$

where $K_0(\cdot)$ is a square-integrable kernel function and $X(\boldsymbol{u})$ is 2-dimensional Brownian motion. There are many interesting features associated with the process specification (6); however, we do not review these here, and refer the reader to Higdon (2002), for a detailed discussion. For application to our specific context, we note that (6) can be used to represent a stationary Gaussian process if and only if it has a spectral density (Xia and Gelfand, 2005). In this case, for a given covariance function $C(\cdot)$, the corresponding kernel can be obtained as the inverse Fourier transform of the square root of the Fourier transform of $C(\cdot)$ (Kern, 2000). In our application, we shall work with a Gaussian kernel having form $K_0(\boldsymbol{u}) = \frac{2}{\sqrt{\pi}\tau_0} \exp\{-2||\boldsymbol{u}||^2/{\tau_0}^2\}$ which, in conjunction with (6), produces a spatial process having an isotropic Gaussian correlation function $C_0(\boldsymbol{u}) = \exp\{-||\boldsymbol{u}||^2/\tau_0^2\}$. For

implementation, (6) is approximated with a discrete sum

$$b_0(\boldsymbol{s}) \approx \sum_{j=1}^{J} K_0(\boldsymbol{u}_j - \boldsymbol{s}) X_{0j}$$

where $\boldsymbol{u}_j,\ j = 1, \ldots, J$ is a set of knots covering the spatial domain, and $X_{0j},\ j = 1, \ldots, J$, is a set of independent and identically distributed mean-zero Gaussian latent variables, with unknown variance that is treated as a parameter in the model. A similar specification is adopted for the spatially-varying slope $b_1(\boldsymbol{s}) \approx \sum_{j=1}^{J} K_1(\boldsymbol{u}_j - \boldsymbol{s}) X_{1j}$ and a bivariate distribution

$$\boldsymbol{X}_j = (X_{0j}, X_{1j})' \overset{iid}{\sim} BVN(\boldsymbol{0}, \boldsymbol{\Sigma}),\ j = 1, \ldots, J, \tag{7}$$

induces correlation between $b_0(\boldsymbol{s})$ and $b_1(\boldsymbol{s})$, with $\Sigma_{12} = 0$ corresponding to independent processes. Choosing $J << n$ results in the required dimension reduction. For our analysis in Section 4, we set $J = 170$ with the resulting grid of knots depicted in Figure 1, panel (b). The bandwidths of the Gaussian kernels, $\tau_l,\ l = 0, 1$, are chosen based on an exploratory analysis, using the fact that the effective range of the corresponding Gaussian correlation function is $R_l = \sqrt{3}\tau_l$, further details are deferred to Section 4.

3.2. Spatial Modeling with Markov Random Fields

An alternative approach for modeling the random effects $\boldsymbol{b} = ((b_0(\boldsymbol{s}_1), b_1(\boldsymbol{s}_1)', \ldots, (b_0(\boldsymbol{s}_n), b_1(\boldsymbol{s}_n)')'$ foregoes the spatial process specification and develops the joint distribution for $\boldsymbol{b}$ based on a Markov random field (MRF). With regularly spaced spatial locations, and with estimation, as opposed to spatial interpolation, being of prime interest, a MRF specification seems reasonable. More specifically, we assume that the random effects are drawn from a bivariate generalization of the Gaussian intrinsic autoregression. Here, the joint distribution is determined through a set of local specifications, based on an $n \times n$ 'neighborhood' matrix $\mathbf{C}$, which we shall take to be binary with $C_{ii} = 0$. The neighborhood matrix encodes a set of conditional independence assumptions underlying the spatial structure of the model. The model for $\boldsymbol{b}$ is then specified through the n conditional distributions

$$\boldsymbol{b}(\boldsymbol{s_i})|\boldsymbol{b}_{-i}, \boldsymbol{\Sigma} \sim BVN(\boldsymbol{\mu_i}, \frac{1}{C_{i+}}\boldsymbol{\Sigma}),\ \ i = 1, \ldots, n, \tag{8}$$

where $\boldsymbol{b}_{-i}$ denotes the vector $\boldsymbol{b}$ with the i^{th} pair $(b_0(\boldsymbol{s_i}), b_1(\boldsymbol{s_i}))'$ removed; $\boldsymbol{\mu_i}$ is defined as a weighted average $\mu_{i_1} = \sum_j \frac{C_{ij}}{C_{i+}} b_0(\boldsymbol{s_j})$, $\mu_{i_2} = \sum_j \frac{C_{ij}}{C_{i+}} b_1(\boldsymbol{s_j})$; and the 2×2 positive definite and symmetric matrix $\boldsymbol{\Sigma}$ is a hyperparameter representing the (conditional) covariance between each pair random effects $(b_0(\boldsymbol{s}), b_1(\boldsymbol{s}))$, with $\Sigma_{12} = 0$ corresponding to independence between the slope and intercept parameters. With these conditional specifications, the corresponding joint distribution for $\boldsymbol{b}$ is obtained as $p(\boldsymbol{b}|\boldsymbol{\Sigma}) \propto \exp\{-1/2\boldsymbol{b}'[(\boldsymbol{D_C} - \boldsymbol{C}) \bigotimes \boldsymbol{\Sigma^{-1}}]\boldsymbol{b}\}$ (Carlin

and Banerjee, 2003) where $\boldsymbol{D_C} = \text{diag}\{C_{1+}, \ldots, C_{n+}\}$. This prior is improper, as the corresponding precision matrix $(\boldsymbol{D_C} - \boldsymbol{C}) \bigotimes \boldsymbol{\Sigma^{-1}}$ is rank deficient. As a result, the constraints $\sum_{i=1}^{n} b_l(\boldsymbol{s_i}) = 0,\ l = 0, 1$ are required for identification of the fixed effect parameters μ_{b0} and μ_{b1} in the skew-t regression (5). We consider several choices of neighborhood matrix $\mathbf{C}$ for our analysis in Section 4, and make model comparisons across the corresponding spatial structures.

3.3. Bayesian Inference and a Missing Data Formulation

Having specified our model at the first level, and distributions for the random effects in (5), our bayesian model specifications are made complete upon assigning prior distributions to the scale, shape and degrees of freedom σ, α, ν of the skew-t distribution; the regression coefficients $\boldsymbol{\beta}$; and the hyperparameters $\sigma_B^2, \sigma_F^2, \Sigma$ controlling variability of the random effects. The prior, for the model based on process convolutions, then factorizes as

$$p(\boldsymbol{\theta}) = \left[\prod_{i=1}^{n_B} p(b_i^{(B)}|\sigma_B^2)\right] \times \left[\prod_{i=1}^{n_F} p(b_i^{(F)}|\sigma_F^2)\right] \times \left[\prod_{i=1}^{J} p(\boldsymbol{X}_j|\boldsymbol{\Sigma})\right]$$

$$\times p(\alpha)p(\sigma)p(\alpha)p(\boldsymbol{\beta})p(\sigma_B^2)p(\sigma_F^2)p(\boldsymbol{\Sigma})$$

and the prior for the MRF model takes a similar form, with the term $\left[\prod_{i=1}^{J} p(\boldsymbol{X}_j|\boldsymbol{\Sigma})\right]$ replaced with the term $p(\boldsymbol{b}|\boldsymbol{\Sigma})$ defined by the conditional specifications (8). We adopt vague prior distributions, in an effort to achieve primarily data driven inference, and these are taken to be conditionally conjugate whenever possible, to facilitate Gibbs updating within an MCMC sampler. The specific forms adopted for the prior specifications are discussed further in Section 4, where the application is considered.

The likelihood function can be obtained as a product

$$L(\boldsymbol{\theta}|\boldsymbol{H}) = \prod_{i=1}^{n}\prod_{j=1}^{m_i} p(H_i(t_{ij})|\mu_i(t_{ij}), \sigma, \alpha, \nu)$$

with each term corresponding to the skew-t density defined in (2). Alternatively, the representation of the skew-t distribution as a Gaussian mixture (4) may be employed, introducing for each i, j, latent variables Z_{ij} and W_{ij}, treated as missing data within an MCMC framework. The complete data likelihood then takes the form

$$L(\boldsymbol{\theta}|\boldsymbol{H}, \boldsymbol{Z}, \boldsymbol{W}) = \prod_{i=1}^{n}\prod_{j=1}^{m_i} p(H_i(t_{ij})|\mu_{H_{ij}}, \sigma_{H_{ij}}^2)p(Z_{ij}|W_{ij})p(W_{ij}|\nu) \tag{9}$$

where $p(H_i(t_{ij})|\mu_{H_{ij}}, \sigma_{H_{ij}}^2)$ is a univariate normal density with mean $\mu_{H_{ij}} = \mu_i(t_{ij}) + \frac{\alpha\sigma Z_{ij}}{\sqrt{1+\alpha^2}}$

and variance $\sigma^2_{H_{ij}} = \frac{\sigma^2}{W_{ij}(1+\alpha^2)}$; $p(Z_{ij}|W_{ij})$ is the density of a $TN_{(0,\infty)}(0, W_{ij}^{-1})$ distribution; and $p(W_{ij}|\nu)$ is the density of a Gamma$(\nu/2, \nu/2)$ distribution. This specification is convenient, as one need only work with a Gaussian likelihood for the observed height data, whilst incorporating additional update steps for the latent variables Z_{ij} and W_{ij}, whose full conditional distributions are easily obtained in closed form, as truncated normal and gamma distributions respectively.

3.4. Model Fitting

We sample from the augmented posterior $p(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{W}|\boldsymbol{H}) \propto L(\boldsymbol{\theta}|\boldsymbol{H}, \boldsymbol{Z}, \boldsymbol{W}) \times p(\boldsymbol{\theta})$ for both the PC and MRF based spatial models, using MCMC samplers based on componentwise transition. Designing an algorithm to fit the PC based model proved the most challenging, and was based on combining Gibbs and Metropolis-Hastings updates in conjunction with hybrid Monte Carlo. We describe the main aspects of this algorithm and then mention modifications required for fitting the MRF model.

In fitting the PC spatial model, we have found that Monte Carlo updating of the latent variables $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_J})$, defined in (7), requires close consideration due to their strong posterior correlations. In particular, random walk updating of these parameters leads to slow movement through the target distribution, despite careful tuning of proposal distributions, reparameterizations, and implementations based on various blocking schemes. As a remedy, we employ a hybrid Monte Carlo algorithm (Gustafson 1997), which is designed to suppress inefficient random walk behavior and promote rapid mixing of the Markov chain. The hybrid algorithm is based on the combination of a stochastic step with a preset number of deterministic steps that represent a discretization of Hamiltonian dynamics. Each iteration of our algorithm involves a block update of $\mathbf{X}$ based on the hybrid procedure. Letting $\pi_{\mathbf{X}}(\cdot)$ denote the density of the target distribution $[\mathbf{X}|\boldsymbol{H}, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\theta}_{-\mathbf{X}}]$, and $\mathbf{X}^*$ the current value in the Markov chain for $\mathbf{X}$, the hybrid update, based on a step size $\delta > 0$, proceeds as follows:

1. *Simulate auxiliary variables* $\mathbf{U}^* \sim MVN_{2J}(\mathbf{0}, \mathbf{I})$.
   Let $\mathbf{X^{(0)}} = \mathbf{X}^*$ *and* $\mathbf{U^{(0)}} = \mathbf{U}^* + \frac{\delta}{2} \nabla \log \pi_{\mathbf{X}}(\mathbf{X}^*)$

2. *For* $l = 1, \ldots, L$, *let*

$$\mathbf{X^{(l)}} = \mathbf{X^{(l-1)}} + \delta\, \mathbf{U^{(l-1)}}$$

$$\mathbf{U^{(l)}} = \mathbf{U^{(l-1)}} + \delta_l \nabla \log \pi_{\mathbf{X}}(\mathbf{X^{(l)}})$$

   where $\delta_l = \delta$ *for* $l < L$ *and* $\delta_L = \frac{\delta}{2}$.

3. *Accept* $\mathbf{X^{(L)}}$ *as the new state for* $\mathbf{X}$ *with probability*

$$p = min\left(\frac{\pi_{\mathbf{X}}(\mathbf{X^{(L)}})}{\pi_{\mathbf{X}}(\mathbf{X}^*)} \exp\left\{-\frac{1}{2}\left(\mathbf{U^{(L)}}'\mathbf{U^{(L)}} - \mathbf{U}^{*\prime}\mathbf{U}^*\right)\right\}, 1\right)$$

else remain in the current state $\mathbf{X}^*$ *with probability* $1-p$.

Each hybrid update requires $L+1$ evaluations of the gradient vector $\nabla \log \pi_{\mathbf{X}}(\cdot)$ and we have found that setting $L = 50$ works well in the current setting. With the mixture representation of the skew-t distribution, based on the complete data likelihood (9), this gradient vector is easily obtained in closed form.

Upon updating $\mathbf{X}$, the sampler visits the remaining nodes, updating with either Gibbs or Metropolis-Hastings steps. Specifically, the random effects $b_l^{(B)}, l = 1, \ldots, n_B$ and $b_l^{(F)}, l = 1, \ldots, n_F$ are sampled directly from their Gaussian full conditional distributions, as are the components of $\boldsymbol{\beta}$, μ_{b0} and μ_{b1}, when Gaussian priors are assigned to these parameters. The variance components σ_B^2 and σ_F^2 are assigned conditionally conjugate inverse-gamma priors, and are sampled from the corresponding inverse-gamma full conditional distributions. In a similar fashion, $\boldsymbol{\Sigma}$ is sampled from an inverse-Wishart full conditional distribution based on an inverse-Wishart prior. The latent variables, Z_{ij} and W_{ij} are also sampled with Gibbs updates, based on their truncated normal and gamma full conditional distributions. Finally, α, σ and ν are individually updated with Metropolis-Hastings steps, based on random walk proposals. Candidate distributions for random walk updates are tuned to yield acceptance rates of between 20 and 50 percent; whereas, δ is chosen so that the hybrid step has an acceptance rate of between 75 and 90 percent. The sampler employed for fitting the MRF based model has a similar structure, with updates for $\boldsymbol{X}$ replaced by updates for $\boldsymbol{b}$. In this case, a hybrid update is not employed; rather, each component of $\boldsymbol{b}$ is drawn from its Gaussian full conditional distribution, which seems to work fine for this model. Software implementation of the sampling algorithm is available in the R programming language from the author upon request.

4. Study of Juvenile White Spruce Growth

Before applying our methodology to the white spruce growth data, we explore the data to shed light on several features that motivate the proposed hierarchical modeling framework. Our data consist of repeated height measurements taken on each of $n = 4330$ trees, with measurements taken at 3, 6 and 10 years after planting, resulting in a total of $4330 \times 3 = 12,990$ observations. In addition, seedling height is recorded at the time of planting, and we shall condition on these initial values, incorporating them as covariates in the regression model. Specifically, seedling height is discretised into three exhaustive categories corresponding to seedlings with initial height less than or equal to 1, 2 and 3 centimeters respectively. Covariate information on the cumulative number of pine weevil attacks on each tree is also recorded at each time point, and the association between this covariate and height growth

is also of interest.

The longitudinal height profiles for each tree depicted in Figure 1, panel (a), suggest a linear trend in height growth over time, with substantial tree-to-tree variability. We begin by fitting a standard linear mixed model for such longitudinal data, incorporating a linear growth curve, with coefficients modeled as independent tree-specific random effects,

$$H_i(t_{ij}) = \beta_1 SH_{2i} + \beta_2 SH_{3i} + \beta_3 N_{ij} + b_{F_i}^{(F)} + b_{B_i}^{(B)} + \mu_{b0} + \mu_{b1} t + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \;\; i = 1, \ldots, 4330, \;\; j = 1, 2, 3$$

where SH_{2i} and SH_{3i} are binary covariates indicating seedling height falling into the second and third categories respectively; N_{ij} is a covariate giving the cumulative number of pine weevil attacks recorded on the i^{th} tree at time t_{ij}; $b_{F_i}^{(F)}$ and $b_{B_i}^{(B)}$ are family and block random effects as defined in (5); and $(b_{0i}, b_{1i})' \overset{iid}{\sim} BVN(\mathbf{0}, \mathbf{\Sigma})$ are the tree-specific effects. This initial model is fit using REML based on standard software for linear mixed models. The normal QQ-plot of the standardized residuals is depicted in Figure 2, panel (a). Clearly, the Gaussian model is inadequate, with the standardized residuals exhibiting both positive skewness and substantial excess kurtosis.

Next, we motivate the need to consider spatially correlated random effects by examining the spatial structure in the raw growth trajectories. To this end, we fit a least-squares line independently at *each* spatial location $H_i(t_{ij}) = b_{0i} + b_{1i} t_{ij}$, obtaining $n = 4330$ exploratory estimates of the site specific slope and intercept parameters. Figure 2, panels (b) and (c) display the empirical semivariogram estimates obtained from these raw values, along with Monte Carlo envelopes obtained by repeated random permutation of the data values on the spatial locations. Positive spatial dependence is evident in both cases, with the intercept exhibiting a correlation length of approximately 10 meters, and the slope exhibiting spatial dependence up to 20 meters of separation. Figure 2, panel (d) displays a scatter plot of these slope and intercept estimates. A strong negative correlation in these raw values $\rho = -0.84$ is evident. Thus, a bivariate spatial model, allowing for this strong negative correlation, seems necessary.

To address these features of the data more formally, we fit the skew-t spatio-temporal regression models proposed in Section 3, incorporating the same regression structure

$$\mu_i(t_{ij}) = \beta_1 SH_{2i} + \beta_2 SH_{3i} + \beta_3 N_{ij} + b_{F_i}^{(F)} + b_{B_i}^{(B)} + \mu_{b0} + \mu_{b1} t_{ij} + b_0(\boldsymbol{s_i}) + b_1(\boldsymbol{s_i}) t_{ij}. \quad (10)$$

For the model based on process convolutions, we cover the spatial domain with a regular grid of knots, based on a spacing of 8 meters between knots. Any knots further than 8 meters from *all* data points are then removed. The resulting set of $J = 170$ knots is depicted in Figure 1, panel (b). The parameters of the Gaussian kernels $K_0(\cdots)$ and $K_1(\cdots)$ are set at

$\tau_0 = 6$ and $\tau_1 = 12$, so that the effective range of the corresponding covariance functions are approximately $R_0 = 10$ meters and $R_1 = 20$ meters, corresponding to the correlation lengths observed in Figure 2, panels (b) and (c).

For fitting the MRF spatial models, we begin with the usual first-order neighborhood structure based on adjacency, so that each tree has either 4, 3 or 2 neighbors, with neighbors being separated by a distance of $d = 1$ meter. We also define three additional neighborhood structures, where trees separated by a distance less than or equal to d are considered neighbors, and we consider neighborhoods based on $d = \sqrt{2}, 10$ and, 20 meters respectively. Finally, as a baseline model for comparison, we consider a model specification incorporating only a fixed effect linear trend, based on setting $b_0(\boldsymbol{s_i}) = b_1(\boldsymbol{s_i}) = 0$ in (10).

In total, we fit six skew-t regression models to these data, using the MCMC sampling schemes described in Section 3.4. Regarding the prior specifications, all fixed effects $\boldsymbol{\beta}, \mu_{b0}, \mu_{b1}$ are assigned conjugate $N(0, 100)$ priors; the block and family variance components σ_B^2 and σ_F^2 are assigned vague inverse-gamma$(.01, .01)$ priors; while $\boldsymbol{\Sigma}^{-1} \sim Wishart(\nu, \mathbf{A})$ where setting $\nu = 2$ and $\mathbf{A} = 0.1 \times \mathbf{I}$ results a relatively vague prior. Finally, the scale, shape and degrees of freedom parameters are assigned $\sigma \sim unif(0, 100)$, $\alpha \sim N(0, 100)$, $\nu \sim unif(.05, 30)$ priors. Our use of weak, conjugate inverse-gamma and Wishart priors for the variance components of the random effects may be criticized (see for example, Gelman, 2006); however, the large size of our dataset suggests little concern with regards to prior sensitivity, in particular since we do not expect negligibly small values for the corresponding variance components.

In fitting all six models, an initial burn-in of $10,000$ iterations seemed more than sufficient for the sampling chains to reach stationarity, and an additional $50,000$ iterations were used to summarize the posterior distribution associated with each model. To compare models, we use the posterior predictive loss criterion proposed by Gelfand and Gosh (1998). Here, we consider estimation of the observed data using the posterior predictive distribution $[\boldsymbol{H}^{(rep)}|\boldsymbol{H}]$, and for squared error loss, we rank the models according to $D = G + P$ where $G = \sum_{i=1}^{n} \sum_{j=1}^{m_{ij}} (H_i(t_{ij}) - \mu_{ij}^{(rep)})^{2,}$ measures the lack of fit, and $P = \sum_{i=1}^{n} \sum_{j}^{m_{ij}} \sigma_{ij}^{2(rep)}$ represents the degree of smoothness; where $\mu_{ij}^{(rep)} = E[H_i(t_{ij})^{(rep)}|\boldsymbol{H}]$ and $\sigma_{ij}^{2(rep)} = Var[H_i(t_{ij})^{(rep)}|\boldsymbol{H}]$ with relatively smaller values of D indicating preferred models.

Table 1 lists Monte Carlo estimates of D for the six models considered, in addition to the corresponding posterior intervals for σ, ν and α. The baseline model, with $b_0(\boldsymbol{s_i}) = b_1(\boldsymbol{s_i}) = 0$ exhibits the poorest performance, indicating that heterogeneity in the growth trajectories is evident, and that models accommodating this feature yield superior performance. Interestingly, the MRF models outperform the PC model by a substantial margin, regardless of neighborhood structure, with the $d = 10$ meter threshold being the most suitable definition of neighbor. This is an interesting result, in particular since the bivariate MRF model

(8) yields a separable bivariate spatial structure; whereas, the PC model allows the slope and intercept processes to exhibit different spatial ranges through choice of kernel functions $K_0(\cdot)$ and $K_1(\cdot)$. To gain further insight into the relatively poorer performance of the PC model, Figure 4, panel (b) compares the posterior mean estimates of $b_1(\boldsymbol{s_i})$, $i = 1 \ldots n$, obtained from the PC model with those obtained from the MRF ($d = 10$) model. There is a clear difference in the estimates obtained from the two models, and the spatial effects estimated from the PC model seem to be 'overshrunk' relative to those obtained from the MRF model. Here, it seems that there is a price to be paid for dimension reduction, and the fully parameterized MRF model is better able to capture the total extent of the tree-to-tree variability. Comparison of the $b_0(\boldsymbol{s_i})$ estimates (not shown) also yields similar results. The performance of the PC model would likely improve if the number of knots J is increased; however, computational demands for the current setting with $J = 170$ knots already exceed those required for fitting the MRF models.

Regardless of the model adopted for the spatial effects, the posterior intervals of ν and α shown in Table 1 all indicate substantial departures from a Gaussian model, with very heavy tails and positive skewness. Compared with the four MRF models, the posterior distributions associated with the baseline model and PC model, favor smaller values of ν and higher values of σ. The relative inability of these models to represent the residual spatial variation leads to heavier tailed errors with greater variability.

Table 2 presents further posterior summaries for the MRF model with neighbors based on the $d = 10$ meter threshold. All covariates included play a significant role in describing variation in height growth. A negative impact of pine weevil attack on tree growth is clearly evident, with greater levels of infection associated with lower height growth overall. In addition, seedling height is positively associated with height growth overall, a result that seems intuitive. The posterior interval of σ_F^2 indicates a non-negligible degree of genetic variability. This is further illustrated in Figure 3 with boxplots, ranked in order of posterior median, summarizing the posterior samples for each of the $n_F = 139$ family effects. Ranking the posterior samples to summarize the family effects in this way is a useful summary, that can aid in directing selections for seed orchards.

In interpreting the posterior distribution of $\boldsymbol{\Sigma}$, we note that the conditional variance in (8) depends on the number of neighbors, through scaling $\boldsymbol{\Sigma}$ by $1/C_{i+}$. We thus scale the estimates of Σ_{11} and Σ_{22} in Table 2 by $1/\overline{C_{i+}}$ where the average number of neighbors with $d = 10$ meters is $\overline{C_{i+}} = 190$. The posterior of the conditional correlation $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ indicates a very strong negative correlation between the slope and intercept parameters. This is not unexpected, given the exploratory estimates depicted in Figure 2, panel (d). However, the magnitude of the correlation in this case implies that $b_1(\boldsymbol{s_i})$ is essentially a linear function of $b_0(\boldsymbol{s_i})$, perhaps further explaining why a separable spatial structure was preferred for modeling the spatial effects.

Finally, turning to the estimated spatial effects, those trees for which $b_1(\boldsymbol{s_i}) > 0$ can be classified as having growth rates above the population average of μ_{b1}. Figure 4, panel (a) marks the locations of those 1699 trees for which $Pr(b_1(\boldsymbol{s_i}) > 0|\boldsymbol{H}) > 0.975$. Growth rates are not uniformly distributed across the plantation, with the subsections located farthest from the origin containing a preponderance of the fastest growing trees. Spatial patterns and residual hot spots such as those revealed in Figure 4 can be useful as they may give clues regarding missing variables that underlie the autocorrelation observed in the map. This can then guide further study with the goal of identifying unknown factors related to height growth.

5. Discussion

We have proposed a mixed longitudinal regression model for spatially correlated growth data, focussing on modeling heavy tails and asymmetry using the skew-t distribution. We have developed methodology that allows for reasonably flexible modeling of space-time variation, while at the same time being computationally tractable for fitting regression models to a large number of non-Gaussian longitudinal trajectories, observed over many spatial locations.

For spatial modeling, both the process convolution and Markov random field approaches pursued here can be refined. In particular, the choice of kernel bandwidths, τ_0 and τ_1 for the former, and neighborhood matrix $\mathbf{C}$ for the latter, were based on an exploratory analysis. An alternative, and preferable approach, would assign priors to these parameters, thus allowing the uncertainty associated with their values to propagate into the variability of the posterior distribution. In this case, informative priors may be required owing to the typically weak identifiability associated with such parameters.

Our model accommodates residual space-time variability through the linear form $f_{\boldsymbol{s_i}}(t) = b_0(\boldsymbol{s_i}) + b_1(\boldsymbol{s_i})t$. A more flexible form for the growth function is easily incorporated, based on $f_{\boldsymbol{s_i}}(t) = \mathbf{b}(\boldsymbol{s})'\mathbf{F}(\mathbf{t})$, where $\mathbf{F}(\mathbf{t})=(F_1(t), \ldots, F_p(t))'$ is a functional basis (polynomials, B-splines, wavelets, etc...) and $\mathbf{b}(\boldsymbol{s})$ is a p-variate spatial process. Nevertheless, the linear structure was sufficient for the specific growth analysis considered here. A related issue is the assumption that the covariate effects, as well as the block and family random effects are constant over time. These terms are thus best interpreted in a time-averaged sense, and specifications incorporating dynamic evolution of these parameters is an interesting avenue for further exploration.

ACKNOWLEDGEMENTS

Research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author thanks Dr. John King of the British Columbia Ministry of Forests for providing the data.

References

Azzalini A, Capitanio, A. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, Series B* **61**: 579-602.

Azzalini A, Capitanio, A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society, Series B* **65**: 367-389.

Banerjee S, Carlin BP, Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC.

Banerjee S, Johnson GA. 2006. Coregionalized Single- and Multi-Resolution Spatially-Varying Growth Curve Modelling with Application to Weed Growth. *Biometrics* **61**: 617-625.

Besag J, York JC, Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**: 1-59.

Calder CA. 2007. Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. *Environmental and Ecological Statistics* **14**: 229-247.

Carlin BP, Banerjee S. 2003. Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In *Bayesian Statistics 7*, (eds.); Bernardo et al., Oxford: Oxford University Press; 45-63.

Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. 2002. *Analysis of Longitudinal Data*, Second Edition, Oxford, U.K.: Oxford University Press.

Fernández C, Steel MFJ. 1998. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359-371.

Gamerman D, Moreire A, Rue H. 2003. Space-varying regression models: specifications and simulation. *Computational statistics and data analysis* **42**: 513-533.

Gelfand AE, Gosh SK. 1998. Model choice: a minimum posterior predictive loss approach. *Biometrika* **85**: 1-11.

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**: 515-533.

Genton MG. 2004. *Skew-elliptical distributions and their applications: A journey beyond normality.* Florida, USA: Chapman and Hall.

Gustafson P. 1997. Large hierarchical bayesian analysis of multivariate survival data. *Biometrics* **53**: 230-242.

Higdon D. 1998. A process-convolution approach to modeling temperatures in the North Atlantic Ocean. *Journal of Ecological and Environmental statistics* **5**: 173-190.

Higdon D. 2002. Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues.* Anderson C, Barnett V, Chatwin PC, El-Shaarawi AH eds. Springer-Verlag: London; 37-56.

Jara A, Quintana FA, Martín ES. 2008. Linear effects mixed models with skew-elliptical distributions: A bayesian approach. *Computational Statistics and Data Analysis* **52**: 5033-5045.

Kern J. 2000. *Bayesian process-convolution approaches to specifying spatial dependence structure.* Ph.D. thesis, Duke University, Durham, NC 27708.

Kim HM, Mallick BK. 2004. A bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference* **120**: 85-101.

Ma Y, Genton MG, Davidian M. 2004. Linear mixed effects models with flexible generalized skew-elliptical random effects. In *Skew-elliptical distributions and their applications: A journey beyond normality.*. Genton MG (ed.); Florida, USA: Chapman and Hall; 339-358.

Xia G, Gelfand AE. 2005. Stationary process approximation for the analysis of large spatial datasets. Technical report, Duke University, Institute of Statistics and Decision Science.
Zimmerman D, Núñez-Antón V. 2001. Parametric modelling of growth curve data: an overview (with discussion). *Test* **10**: 1-73.

Zhou W, Sansó B. 2007. Statistical inference for atmospheric transport models using process convolutions. *Environmetrics* **19**: 87-101.

List of Tables

List of Figures

Table 1: Monte Carlo estimates of the posterior predictive loss criterion D, and 95% equal tail posterior credible intervals for ν, α and σ for the six skew-t models considered.

Model	D^*	ν	α	σ
No Spatial Effects	30,877,686	(2.19, 2.46)	(0.06, 0.22)	(18.21, 19.31)
Process Convolution, $J = 170$	12,626,174	(2.25, 2.54)	(0.41,1.63)	(16.93, 17.94)
Markov Random Field, $d = 1$	418,897	(3.07, 3.67)	(0.95, 1.39)	(8.87, 10.09)
Markov Random Field, $d = \sqrt{2}$	428,482	(3.06, 3.68)	(0.96,1.40)	(8.88, 10.14)
Markov Random Field, $d = 10$	109,181	(3.21, 3.84)	(1.18, 1.57)	(9.60, 10.67)
Markov Random Field, $d = 20$	300,752	(3.21, 3.74)	(1.14, 1.46)	(9.45, 10.43)

* Values of D after subtracting $2,000,000$

Table 2: Posterior mean and 95% equal tail credible intervals for select parameters of the MRF model with $d = 10$ meters.

Parameter	Posterior Mean	95% Credible Interval
μ_{b_0}	-17.03	(-18.07, -15.95)
μ_{b_1}	20.98	(20.88, 21.06)
SH_{2i}	5.13	(4.51, 5.70)
SH_{3i}	10.37	(9.42, 11.35)
N_{ij}	-16.91	(-17.58, -16.13)
σ_B^2	1.22	(0.22, 3.98)
σ_F^2	2.87	(1.73, 4.22)
$\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$	-0.99	(-1.00, -0.99)
$\Sigma_{11}/\overline{C_{i+}}$	161.26	(152.84, 169.89)
$\Sigma_{22}/\overline{C_{i+}}$	35.90	(34.36, 37.49)

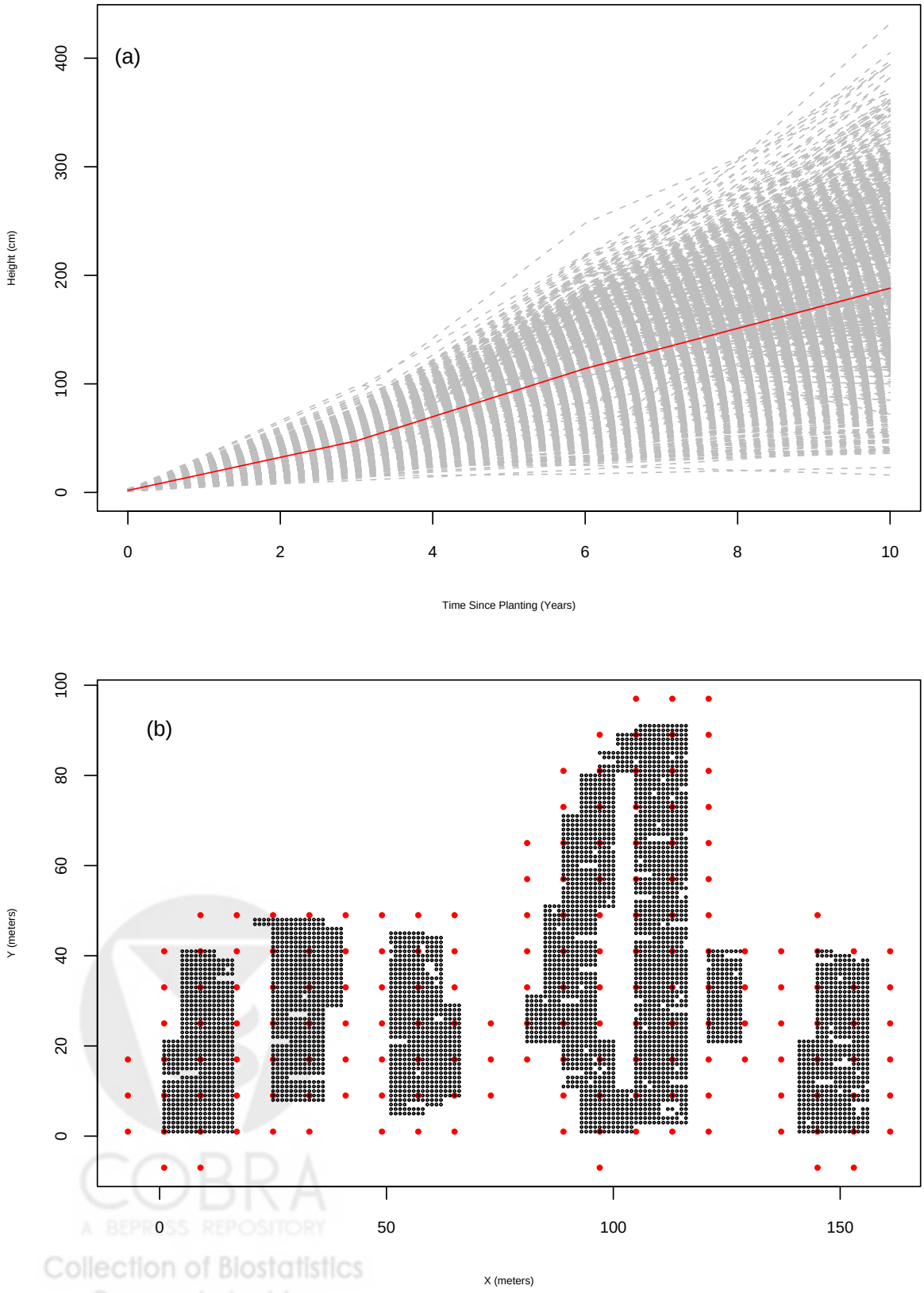

Figure 1: Panel (a) depicts the individual height profiles for all $n = 4330$ trees along with the overall mean; panel (b) marks the locations of trees within the plantation with the knot locations (larger circles) for the discrete kernel convolution superimposed.

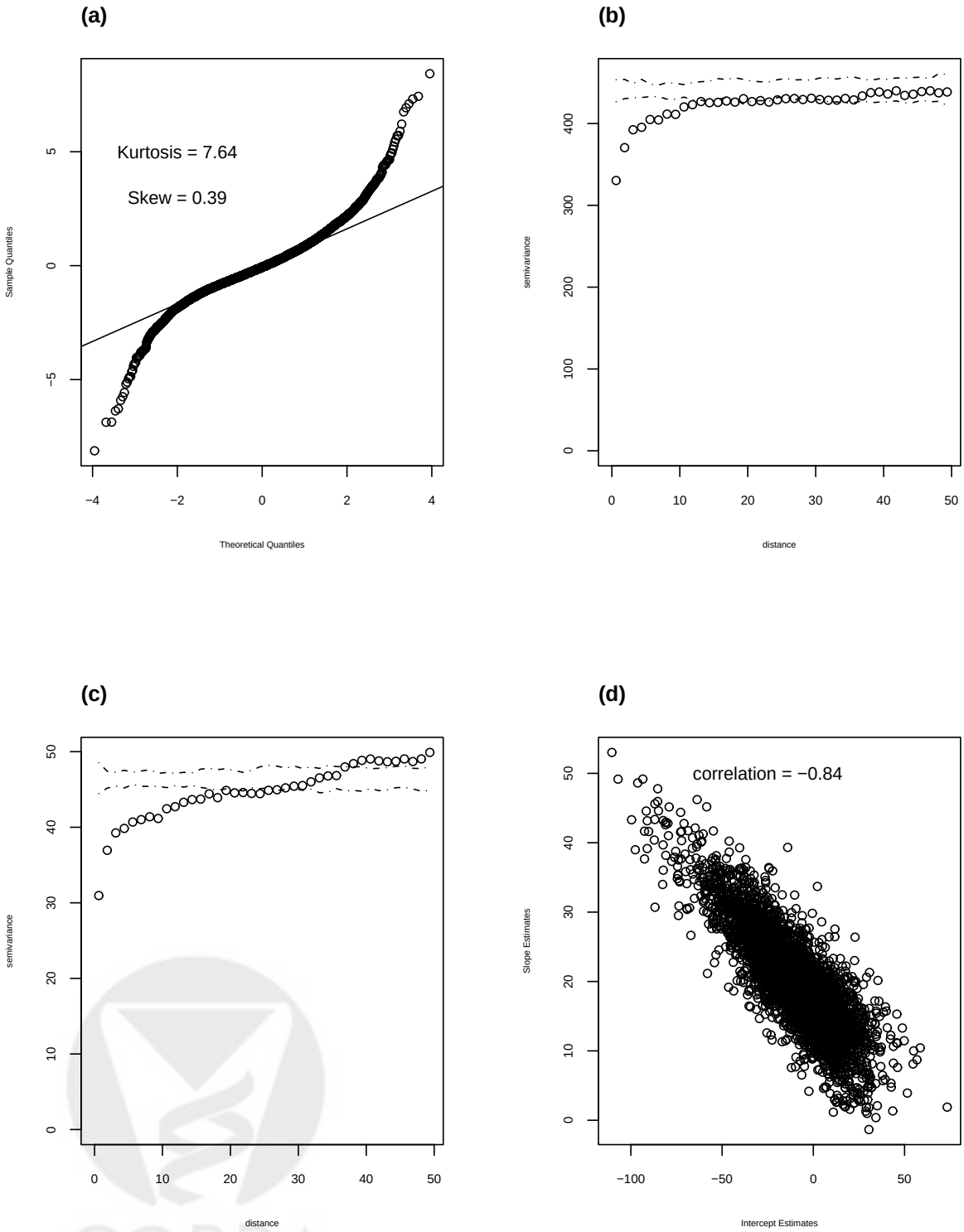

Figure 2: Exploratory analysis of the white spruce growth data: (a) normal QQ-plot of the standardized residuals from a Gaussian linear mixed model; (b) empirical semivariogram of the raw site-specific intercept estimates; (c) empirical semivariogram of the raw site-specific slope estimates; (d) scatter-plot illustrating strong negative correlation between the raw slope and intercept estimates.

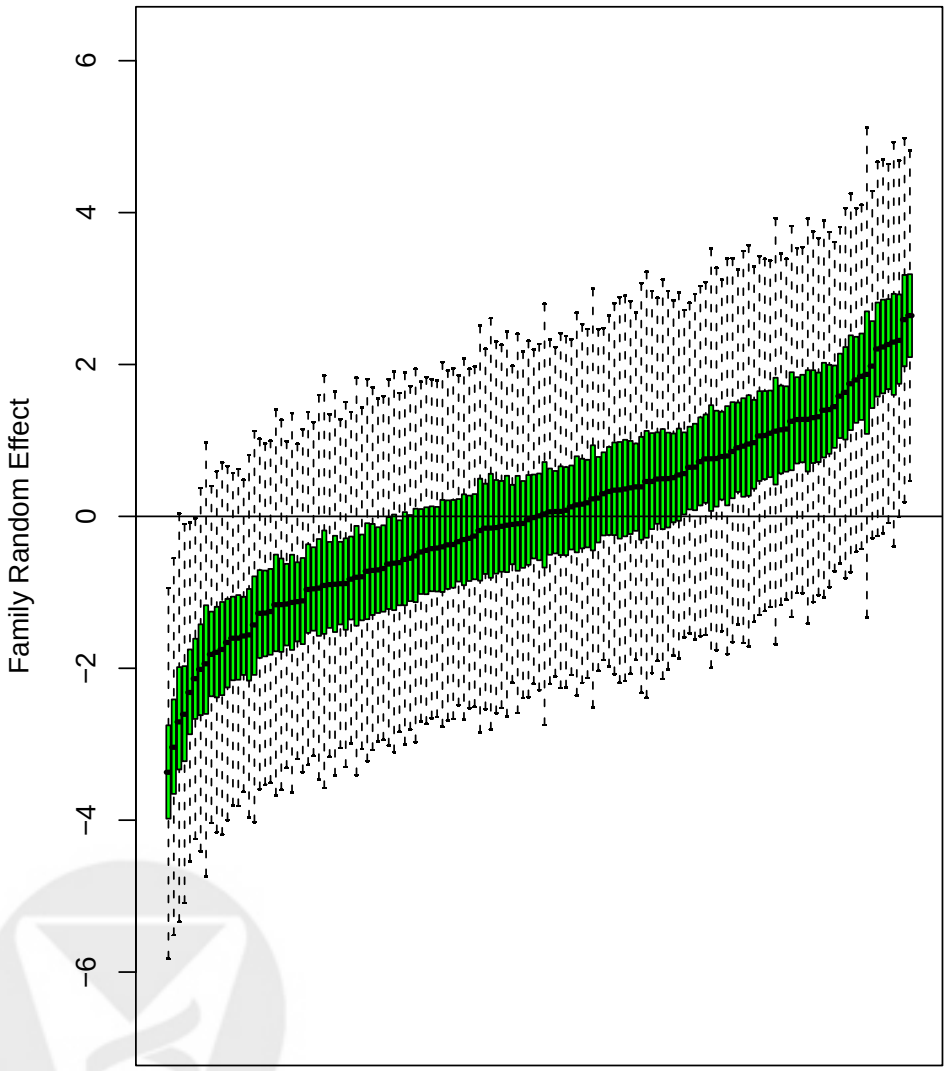

Figure 3: Boxplots (arranged in increasing order by posterior median) illustrating the posterior distribution of the family random effects $b_l^{(F)}$, $l = 1, \ldots, 139$.

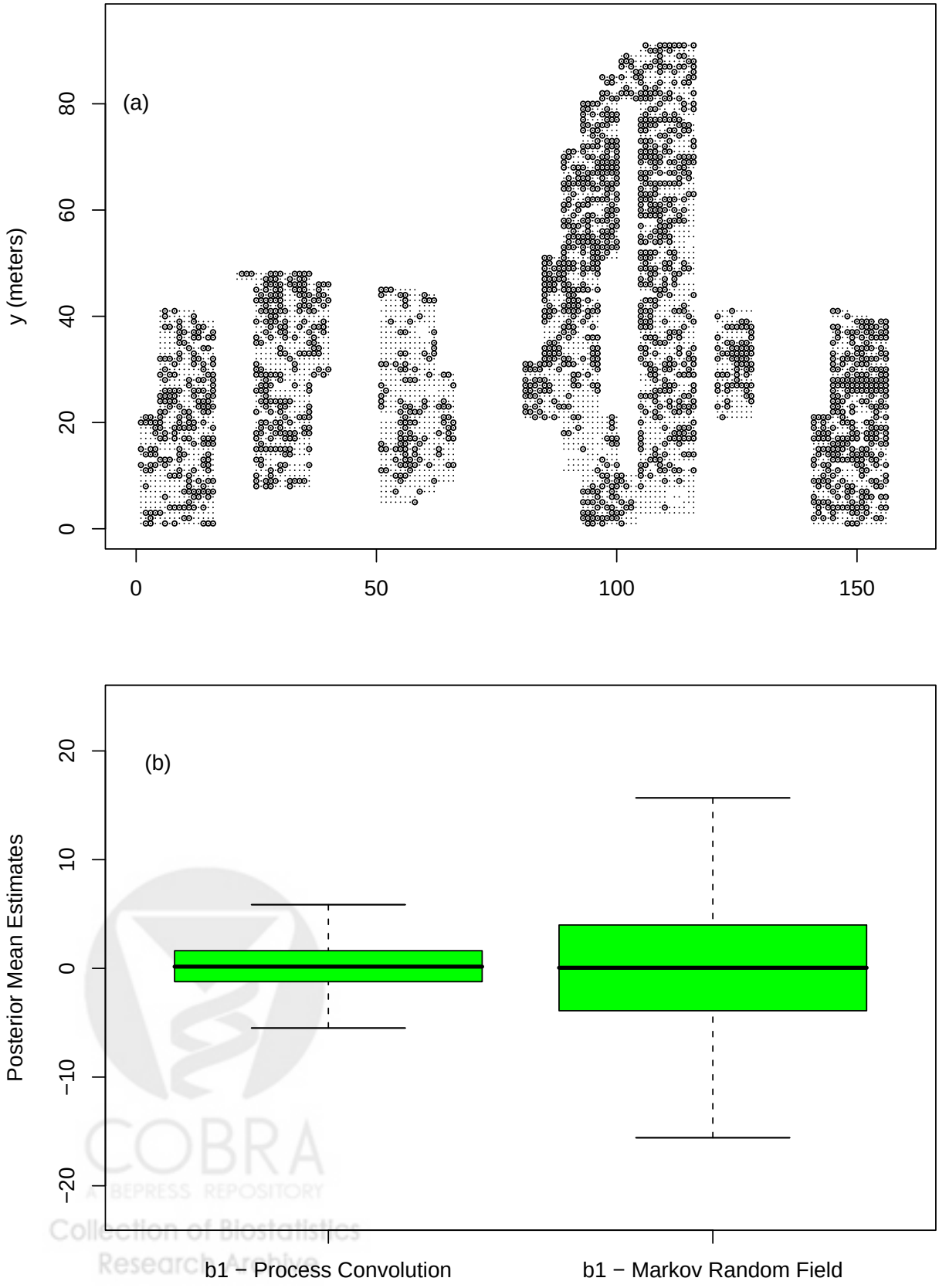

Figure 4: Panel (a) marks the locations of those 1699 trees for which $Pr(b_1(\boldsymbol{s_i}) > 0|\boldsymbol{H}) > 0.975$; panel (b) compares posterior mean estimates of $b_1(\boldsymbol{s_i})$, $i = 1, \ldots, 4330$ obtained from the PC and MRF models.