

# Hierarchical Spatial Modeling of Additive and Dominance Genetic Variance for Large Spatial Trial Datasets

Andrew O. Finley,<sup>1,\*</sup> Sudipto Banerjee,<sup>2</sup> Patrik Waldmann,<sup>3</sup> and Tore Ericsson<sup>3,4</sup>

<sup>1</sup>Department of Forestry and Department of Geography, Natural Resources Building, Michigan State University, East Lansing, Michigan 48824-1222, U.S.A.

<sup>2</sup>Department of Biostatistics, School of Public Health, University of Minnesota, A 460 Mayo Building, MMC 303, 420 Delaware Street, Minneapolis, Minnesota 55455, U.S.A.

<sup>3</sup>Department of Forest Genetics and Plant Physiology, Swedish Agricultural University, 901 83 Umeå, Sweden

<sup>4</sup>Skogforsk, Box 3, 918 21 Sävar, Sweden

\**email:* finleya@msu.edu

**SUMMARY.** This article expands upon recent interest in Bayesian hierarchical models in quantitative genetics by developing spatial process models for inference on additive and dominance genetic variance within the context of large spatially referenced trial datasets. Direct application of such models to large spatial datasets are, however, computationally infeasible because of cubic-order matrix algorithms involved in estimation. The situation is even worse in Markov chain Monte Carlo (MCMC) contexts where such computations are performed for several iterations. Here, we discuss approaches that help obviate these hurdles without sacrificing the richness in modeling. For genetic effects, we demonstrate how an initial spectral decomposition of the relationship matrices negate the expensive matrix inversions required in previously proposed MCMC methods. For spatial effects, we outline two approaches for circumventing the prohibitively expensive matrix decompositions: the first leverages analytical results from Ornstein–Uhlenbeck processes that yield computationally efficient tridiagonal structures, whereas the second derives a modified *predictive process* model from the original model by projecting its realizations to a lower-dimensional subspace, thereby reducing the computational burden. We illustrate the proposed methods using a synthetic dataset with additive, dominance, genetic effects and anisotropic spatial residuals, and a large dataset from a Scots pine (*Pinus sylvestris* L.) progeny study conducted in northern Sweden. Our approaches enable us to provide a comprehensive analysis of this large trial, which amply demonstrates that, in addition to violating basic assumptions of the linear model, ignoring spatial effects can result in downwardly biased measures of heritability.

**KEY WORDS:** Bayesian inference; Genetic variance; Markov chain Monte Carlo; Ornstein–Uhlenbeck process; Spatial predictive process; Spatial process.

## 1. Introduction

Hierarchical random effects models implemented through Markov chain Monte Carlo (MCMC) methods have recently gained popularity within quantitative genetics (see, e.g., Sorensen and Gianola, 2002; Waldmann and Ericsson, 2006, and references therein). Both single-component and blocked Gibbs samplers have been proposed. However, the single-component samplers suffer from slow mixing and the blocked samplers often involve matrix decompositions whose complexity increases as  $O(n^3)$  in the number of observations,  $n$ , at every iteration of the MCMC algorithm. These computationally intensive matrix calculations limit inference to small sample sizes, which might lead to unreliable results and inconclusive evidence.

Quantitative genetics studies the inheritance of polygenic traits, focusing upon estimation of additive genetic variance,  $\sigma_a^2$ , and the heritability  $h^2 = \sigma_a^2/\sigma_{Tot}^2$ , where the denominator represents the total genetic and unexplained variation. A high heritability should result in a larger selection response, i.e., a higher probability for genetic gain in future generations. Genetic variance may be further partitioned into additive and

nonadditive components (Lynch and Walsh, 1998). Given a polygenic variable of interest and individual pedigree information coupled with an experimental design with collateral relations, mixed effects models can provide inference of all genetic variances (Henderson, 1985) as the additive, dominance, and epistatic genetic effects are uncorrelated in unselected, non-inbred populations in linkage equilibrium (Cockerham, 1954). Given the dominance genetic variance,  $\sigma_d^2$ , the dominance ratio,  $d^2 = \sigma_d^2/\sigma_{Tot}^2$ , is often estimated, but currently lacks direct practical interpretation.

A feature common to genetic trials in forestry and agriculture is the presence of systematic heterogeneity among observational units. If this spatial heterogeneity is neglected in the model, estimates of the genetic parameters might be biased (Dutkowski et al., 2002; Cappa and Cantet, 2007). For spatially arranged observational units, heterogeneity typically results from small-scale environmental conditions (e.g., soil characteristics, microclimates, light availability, etc.) that might cause correlation among units as a function of distance and/or direction. Although traditional randomized designs have sought to negate any confounding spatial correlation,

recent interest has shifted to models accounting for spatial correlation that, most often, incorporate spatial correlation through differencing or by constructing covariates using residuals from neighboring units (see e.g., Zimmerman and Harville, 1991; Cullis et al., 1998, and references therein).

A notable exception is Zimmerman and Harville (1991), who explore classical geostatistical methods for including spatial covariances in their models. We too adopt a geostatistical approach, but focus upon settings where the number of locations is too large for efficiently estimating hierarchical spatial models. We offer computationally efficient methods for Bayesian estimation of genetic variance components in the individual-based linear mixed model with or without inclusion of spatial random effects. For genetic effects, we employ a spectral decomposition of the relationship matrices to obviate the expensive matrix inversions required in previously proposed MCMC samplers. We restrict ourselves to additive and dominance effects to avoid identifiability issues arising through epistatic effects.

For spatial effects, we explore two approaches for circumventing the expensive matrix decompositions. The first constructs spatial processes from products of special correlation functions that yield sparse tridiagonal structures. This method has been investigated for likelihood based methods by Martin (1990); here, we adapt it to a hierarchical Bayesian framework with missing data. Our second method is less familiar, especially in genetic trials: we apply a predictive process model (Banerjee et al., 2008) that dwells in a lower-dimensional subspace with lower computational burden. A shortcoming of the original formulation of the predictive process is that it induces a positive bias in the nonspatial error term of the models. To address this shortcoming, we proposed a modified predictive process that effectively removes this bias. In Section 2, we develop the models for the data and detail a method to aid in model selection. In Section 3, we outline the proposed MCMC sampler and discuss implementation for the genetic and spatial effects. A synthetic data analysis and the Scots pine analyses are presented in Section 4. In Web Appendix B, we provide additional comparative analysis between the proposed spatial models over a subset of the Scots pine trial site. In Section 5, we conclude the article with a brief discussion.

**2. Models for Genetic and Spatial Effects**

Henderson (1985) showed how a standard linear mixed model can be used to estimate the *additive* and *dominance* genetic variance components for an individual, given as

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + a_i + d_i + \epsilon_i, \tag{1}$$

where  $Y_i$  is the  $i$ th individual’s phenotypic trait of interest,  $\mathbf{x}_i$  is its  $p \times 1$  vector of fixed effects regressors,  $\boldsymbol{\beta}$  is the corresponding  $p \times 1$  slope vector, and  $a_i$  and  $d_i$  are individual specific additive and dominance *random effects*, respectively. Model measurement error is captured by the independent and normally distributed  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \tau^2)$ . For a collection of individuals, we construct  $\mathbf{Y} = [Y_i]_{i=1}^n$  and  $\mathbf{X} = [\mathbf{x}_i^T]_{i=1}^n$ . We assume a multivariate normal,  $MVN(\boldsymbol{\mu}_\beta, \Sigma_\beta)$ , prior distribution for  $\boldsymbol{\beta}$ , whereas the  $n \times 1$  vectors of random effects  $\mathbf{a} = [a_i]_{i=1}^n$  and  $\mathbf{d} = [d_i]_{i=1}^n$  are distributed as  $MVN(\boldsymbol{\mu}_a, \sigma_a^2 \mathbf{A})$  and  $MVN(\boldsymbol{\mu}_d,$

$\sigma_d^2 \mathbf{D})$ , respectively, where  $\boldsymbol{\mu}_a$  and  $\boldsymbol{\mu}_d$  are customarily zero vectors,  $\mathbf{A}$  is the fixed additive relationship matrix, and  $\mathbf{D}$  is the fixed dominance relationship matrix. Lynch and Walsh (1998) detail calculation of  $\mathbf{A}$  and  $\mathbf{D}$  given pedigree information.

Model (1) is adequate in the absence of extraneous variation, beyond what is explained by possible covariates and the additive and dominance random effects. However, if individuals are observed at locations  $\mathbf{s} \in D \subseteq \mathbb{R}^2$ , then we might expect similar phenotypic responses in proximate locations, possibly resulting from similar environmental conditions. Ignoring this spatial dependence may result in falsely precise estimates of  $\boldsymbol{\beta}$  and erroneous variance estimates and suggests adding an additional random effect accounting for spatial association as a decreasing function of distance between locations.

With observations arising over a set of locations, say  $\mathcal{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$ , for each site  $\mathbf{s}_i$  we add a random spatial effect to model (1) for capturing any spatial association,

$$Y(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + a_i + d_i + w(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{2}$$

where the  $\mathbf{s}_i$  reminds us that an individual’s phenotypic response is now associated with a given location. The  $w(\mathbf{s}_i)$ ’s provide local adjustment (with structured dependence) to the mean and are assumed to arise as realizations of a univariate Gaussian process (GP), written  $w(\mathbf{s}) \sim GP\{0, \sigma_w^2 \rho(\cdot; \boldsymbol{\theta})\}$  with mean 0, constant process variance  $\sigma_w^2$ , and correlation function  $\rho(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  are parameters specifying correlation behavior. For a collection of  $n$  locations,  $\mathbf{w} \sim MVN\{\mathbf{0}, \sigma_w^2 R(\boldsymbol{\theta})\}$ , where  $R(\boldsymbol{\theta}) = [\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})]_{i,j=1}^n$  is the correlation matrix. Under assumptions of *stationarity*, the correlation depends upon the separation,  $\rho(\mathbf{s}_i - \mathbf{s}_j; \boldsymbol{\theta})$ , whereas under *isotropy* it depends upon the distance and we write  $\rho(d_{ij}; \boldsymbol{\theta})$ , where  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ . In equation (2)  $w(\mathbf{s}_i)$  often captures the effect of unmeasured or unobserved spatial covariates, whereas  $\epsilon(\mathbf{s}_i)$  models measurement error (the *nugget*; Cressie, 1993) or microscale variation.

The nonstationary Matérn correlation function (Paciorek and Schervish, 2006),

$$\begin{aligned} \rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_{ij}) &= \frac{1}{\Gamma(\nu)2^{\nu-1}} |\Sigma(\mathbf{s}_i)|^{1/4} |\Sigma(\mathbf{s}_j)|^{1/4} \\ &\times \left| \frac{\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j)}{2} \right|^{-1/2} (2\sqrt{\nu d_{ij}})^\nu \kappa_\nu(2\sqrt{\nu d_{ij}}), \end{aligned}$$

is an especially versatile choice for  $\rho(\cdot; \boldsymbol{\theta})$ . Here  $\boldsymbol{\theta}_{ij} = \{\nu, \Sigma(\mathbf{s}_i), \Sigma(\mathbf{s}_j)\}$  varies across the domain (nonstationary),  $\Gamma(\nu)$  is the gamma function,  $\kappa_\nu$  is a second kind of Bessel function, and  $d_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T \{(\Sigma(\mathbf{s}_i) + \Sigma(\mathbf{s}_j))/2\}^{-1} (\mathbf{s}_i - \mathbf{s}_j)$  is the Mahalanobis distance between  $\mathbf{s}_i$  and  $\mathbf{s}_j$ . When  $\Sigma(\mathbf{s}) = \Sigma$  is constant over space, we obtain  $\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}) = \{1/\Gamma(\nu)2^{\nu-1}\} (2\sqrt{\nu d_{ij}})^\nu \kappa_\nu(2\sqrt{\nu d_{ij}})$ , where  $d_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T \Sigma^{-1} (\mathbf{s}_i - \mathbf{s}_j)$ . This produces a stationary but anisotropic process, which we further parameterize as  $\Sigma = G(\psi)\Lambda^2 G^T(\psi)$ , where  $G(\psi)$  is a rotation matrix with angle  $\psi$  and  $\Lambda$  is the diagonal matrix with positive diagonal elements,  $\lambda$ ’s. Thus,  $\boldsymbol{\theta} = (\nu, \psi, \Lambda)$ . The parameter  $\nu$  controls the smoothness whereas the rate of spatial decay is controlled by the  $\lambda^2$ ’s. Finally, a stationary and isotropic process results with  $\Sigma = I$ , where

$\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}) = (1/\Gamma(\nu)2^{\nu-1})(\sqrt{\phi d_{ij}})^{\nu} \kappa_{\nu}(\sqrt{\phi d_{ij}})$  with  $\Sigma^{-1} = I$  and where  $\boldsymbol{\theta} = (\nu, \phi)$  with  $\phi$  controlling spatial decay. This is further simplified by fixing  $\nu = 0.5$ , yielding the familiar exponential correlation  $\rho(\mathbf{s}_i, \mathbf{s}_j; \phi) = \exp(-\phi d_{i,j})$ .

### 2.1 Spatial Process Models for Large Datasets

For a large number of spatial locations, fitting customary geostatistical models becomes prohibitive with necessary matrix factorizations of cubic-order complexities. Fitting the model in equation (2) over  $n$  observations in  $\mathcal{S}$  involves evaluating the data likelihood  $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \Sigma_{\mathbf{Y}})$ , with  $\mathbf{Y} = [Y(\mathbf{s}_i)]_{i=1}^n$ ,  $X = [\mathbf{x}^T(\mathbf{s}_i)]_{i=1}^n$  and  $\Sigma_{\mathbf{Y}} = \sigma_a^2 \mathbf{A} + \sigma_d^2 \mathbf{D} + \sigma_w^2 R(\boldsymbol{\theta}) + \tau^2 I_n$ . Under a Bayesian setup, we assign prior distributions to  $\Omega = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_a^2, \sigma_d^2, \sigma_w^2, \tau^2)$  and inference proceeds by sampling from the posterior distribution  $p(\Omega | \mathbf{Y})$ , whereas prediction at an arbitrary site  $\mathbf{s}_0$  samples  $p(Y(\mathbf{s}_0) | \mathbf{Y})$  one-for-one with posterior draws of  $\Omega$  by composition (Banerjee, Carlin, and Gelfand, 2004). Evidently, both estimation and prediction require evaluating the  $n \times n$  matrix  $\Sigma_{\mathbf{Y}}^{-1}$ , which becomes prohibitive for large  $n$ . This is referred to as the ‘‘big  $n$  problem’’ in spatial statistics and has received considerable recent attention. Existing approaches include kernel-convolution approximations and low-rank kriging methods using splines (Higdon, 2002; Kamman and Wand, 2003; Ver Hoef, Cressie, and Barry, 2004; Cressie and Johannesson, 2008), computationally efficient likelihood approximations either in the spectral domain (Whittle, 1954; Stein, 1999; Fuentes, 2002; Paciorek, 2007), or with appropriate conditional distributions (e.g., Vecchia, 1988; Stein, Chi, and Welty, 2004) and sparse approximations using Markov random field models (Rue and Tjelmeland, 2002; Rue and Held, 2006).

For spatial progeny trials, Martin (1990) proposed tensor product spatial models using univariate stochastic processes along each coordinate. This method leads to *separable* spatial correlation structures forming valid spatial processes (see, e.g., Stein, 1999, Section 2.11) and can indeed lead to computationally feasible models. However, this approach is convenient only for locations on rectangular grids and also imposes modeling limitations. To obviate such limitations, we propose employing a *predictive process* model derived from the original process, but first we offer a brief review of the tensor product models.

**2.1.1 Modeling with Ornstein–Uhlenbeck (OU) processes.** Consider a Gaussian process  $w(s) \sim \text{GP}\{0, \sigma^2 \rho(\cdot; \phi)\}$  on the real line  $\mathbb{R}^1$ , with  $\rho(s - s'; \phi) = \exp(-\phi|s - s'|)$ . This is known as the OU process whose realizations over  $s_1 < \dots < s_n$  yield a tridiagonal inverse correlation matrix with  $[R^{-1}(\phi)]_{ij} = 0$  if  $|i - j| > 1$ , the super- and subdiagonal elements  $[R^{-1}(\phi)]_{ij} = -\rho_{i,j}^2(\phi)/(1 - \rho_{i,j}^2(\phi))$  when  $|i - j| = 1$ , where  $\rho_{i,j}(\phi) = \exp(-\phi|s_i - s_j|)$ , and with diagonal elements  $[R^{-1}(\phi)]_{ii}$  as

$$\begin{aligned}
 & 1 + I[i = 1] \frac{\rho_{i+1,i}^2(\phi)}{1 - \rho_{i+1,i}^2(\phi)} + I[i = n] \frac{\rho_{i,i-1}^2(\phi)}{1 - \rho_{i,i-1}^2(\phi)} \\
 & + I[1 < i < n] \sum_{j=0}^1 \frac{\rho_{i-j,i-j+1}^2(\phi)}{1 - \rho_{i-j,i-j+1}^2(\phi)}.
 \end{aligned}$$

The tridiagonal inverse implies a Markovian structure for the  $w(s_i)$ 's because  $[R^{-1}(\phi)]_{ij} = 0$  if and only if  $w(s_i)$  and  $w(s_j)$

are conditionally independent given the remaining  $w(s)$ 's. Therefore, the conditional distribution of  $w(s_k)$  given the remaining variables will depend only upon its two ‘‘neighbors’’  $w(s_{k-1})$  and  $w(s_{k+1})$ .

In the spatial setting we model  $w(\mathbf{s})$  on  $\mathbb{R}^2$  by assuming its correlation function to be separable over its coordinates. Denoting  $\mathbf{s} = (s_1, s_2)$  and  $\mathbf{s}' = (s'_1, s'_2)$ , we construct  $\rho(\mathbf{s}, \mathbf{s}') = \exp(-\phi_1 |s_1 - s'_1|) \exp(-\phi_2 |s_2 - s'_2|)$ , which results in anisotropic processes whenever  $\phi_1 \neq \phi_2$ . With  $n = n_1 n_2$  locations in  $\mathcal{S}$  falling on a two-dimensional rectangular grid, say  $n_1$  along the first coordinate (rows) and  $n_2$  along the second (columns), the  $n_1 n_2 \times n_1 n_2$  inverse correlation matrix becomes a Kronecker (tensor) product of 2 tridiagonal matrices, i.e.,  $R^{-1}(\phi_1, \phi_2) = R_1^{-1}(\phi_1) \otimes R_2^{-1}(\phi_2)$ , where  $R_1(\phi_1)$  and  $R_2(\phi_2)$  are the  $n_1 \times n_1$  and  $n_2 \times n_2$  correlation matrices arising from OU processes along the  $s_1$  and  $s_2$  axes, respectively. Tridiagonal matrix decompositions enjoy linear time computational complexity, rendering these computations feasible (also see Section 3).

Although the OU models discussed above arise from univariate  $AR(1)$  processes, this is indeed a valid method for constructing spatial process models from univariate processes. See Martin (1990) for other models arising from  $ARMA(p, q)$  processes. However, its computational feasibility is compromised with more general correlation structures and also with nongridded locations. The latter can be relaxed using imputations to a grid (e.g., Fuentes, 2007; Paciorek, 2007) but may introduce unquantifiable uncertainty in the process. Our second approach, discussed below, does not suffer from these drawbacks.

**2.1.2 Predictive process models.** Recently Banerjee et al. (2008) proposed a class of knot-based spatial process models for large spatial datasets. These models consider a set of ‘‘knots’’  $\mathcal{S}^* = (\mathbf{s}_1^*, \dots, \mathbf{s}_m^*)$  with  $m \ll n$ , which may or may not form a subset of the entire collection of observed locations in  $\mathcal{S}$ . The Gaussian process  $w(\mathbf{s})$  yields an  $m$ -vector of realizations over the knots, say  $\mathbf{w}^* = [w(\mathbf{s}_i^*)]_{i=1}^m \sim \text{MVN}\{\mathbf{0}, \sigma_w^2 R^*(\boldsymbol{\theta})\}$  where  $R^*(\boldsymbol{\theta}) = [\rho(\mathbf{s}_i^*, \mathbf{s}_j^*; \boldsymbol{\theta})]_{i,j=1}^m$  is the corresponding  $m \times m$  dispersion matrix. Spatial interpolation (or ‘‘kriging’’) at a generic site  $\mathbf{s}$

$$\tilde{w}(\mathbf{s}) = E\{w(\mathbf{s}) | \mathbf{w}^*\} = \mathbf{r}^T(\mathbf{s}; \boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta}) \mathbf{w}^*, \quad (3)$$

yields a spatial process  $\tilde{w}(\mathbf{s}) \sim \text{GP}\{0, \sigma_w^2 \tilde{\rho}(\cdot)\}$ , where  $\tilde{\rho}(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}) = \mathbf{r}^T(\mathbf{s}; \boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta}) \mathbf{r}(\mathbf{s}', \boldsymbol{\theta})$  and  $\mathbf{r}(\mathbf{s}; \boldsymbol{\theta}) = [\rho(\mathbf{s}, \mathbf{s}_j^*; \boldsymbol{\theta})]_{j=1}^m$ . We refer to  $\tilde{w}(\mathbf{s})$  as the *predictive process* derived from the *parent process*  $w(\mathbf{s})$ . The predictive process is a spatially adaptive linear transformation of the realizations of  $w(\mathbf{s})$  over  $\mathcal{S}^*$  with  $\mathbf{r}^T(\mathbf{s}; \boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta})$  comprising the coefficients of the transformation. This also implies that  $\tilde{w}(\mathbf{s})$  is nonstationary, even though  $w(\mathbf{s})$  is, allowing the model to adapt better to fit the data. Dimension reduction occurs as follows: because  $\tilde{\mathbf{w}} = [\tilde{w}(\mathbf{s}_i)]_{i=1}^n = \mathbf{r}^{*T}(\boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta})$ , where  $\mathbf{r}^{*T}(\boldsymbol{\theta}) = [\mathbf{r}^{*T}(\mathbf{s}_i)]_{i=1}^n$  is  $n \times m$ , the likelihood computations now involve computing the  $m \times m$  matrix  $R^{*-1}(\boldsymbol{\theta})$ , where  $m$  is chosen to be much smaller than  $n$ . Unlike other knot-based methods, the predictive process does not introduce additional parameters (e.g., Higdon, 2002) nor involves projecting data onto a grid (e.g., Paciorek, 2007) while enjoying attractive theoretical properties that justify its use as a *best approximation* for the parent process. For example,  $\tilde{w}(\mathbf{s})$  is an orthogonal projection of  $w(\mathbf{s})$

on an appropriate linear subspace (e.g., Stein, 1999) minimizing  $E[\{w(\mathbf{s}) - f(\mathbf{w}^*)\}^2 | \mathbf{w}^*]$  over all real-valued functions  $f(\mathbf{w}^*)$ .

2.1.3 *Modified predictive process and implementation.* A drawback of the predictive process model proposed by Banerjee et al. (2008) is that it systematically underestimates the variance of the parent process  $w(\mathbf{s})$  at any location  $\mathbf{s}$ . This follows from the veracity of the following inequality

$$\begin{aligned} 0 \leq \text{var}\{w(\mathbf{s}) | \mathbf{w}^*\} &= \sigma_w^2 \{1 - \mathbf{r}^T(\mathbf{s}, \boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta}) \mathbf{r}(\mathbf{s}, \boldsymbol{\theta})\} \\ &= \sigma_w^2 - \text{var}\{\tilde{w}(\mathbf{s})\}. \end{aligned}$$

This can become problematic in estimating different variance components as the underestimation of the spatial variance  $\sigma_w^2$  often leads to overestimation of the other variance components, e.g.,  $\sigma_a^2$ ,  $\sigma_d^2$ , and  $\tau^2$ . Hence, we investigate a *modified predictive process*, defined as  $\tilde{w}_\epsilon(\mathbf{s}) = \tilde{w}(\mathbf{s}) + \tilde{\epsilon}(\mathbf{s})$ , where  $\tilde{\epsilon}(\mathbf{s}) \stackrel{\text{indep}}{\sim} N[0, \sigma_w^2 \{1 - \mathbf{r}^T(\mathbf{s}, \boldsymbol{\theta}) R^{*-1}(\boldsymbol{\theta}) \mathbf{r}(\mathbf{s}, \boldsymbol{\theta})\}]$  is an independent process but with a spatially adaptive variance. Indeed, we now have  $\text{var}\{\tilde{w}_\epsilon(\mathbf{s})\} = \text{var}\{w(\mathbf{s})\}$  as desired. Furthermore,  $E\{\tilde{w}_\epsilon(\mathbf{s}) | \mathbf{w}^*\} = \tilde{w}(\mathbf{s})$  so  $\tilde{w}_\epsilon(\mathbf{s})$  inherits the attractive approximation properties of  $\tilde{w}(\mathbf{s})$ . Note that no new parameters are introduced, hence the adjustment is identifiable, and the computational benefits of  $\tilde{w}(\mathbf{s})$  are retained; we now work with the realizations  $\tilde{\mathbf{w}}_\epsilon = \tilde{\mathbf{w}} + \tilde{\epsilon}$ , where  $\tilde{\epsilon} = [\tilde{\epsilon}(\mathbf{s}_i)]_{i=1}^n$  has a diagonal variance-covariance matrix.

We conclude with some remarks on knot selection—for details see Banerjee et al. (2008) and the references therein. Perhaps the most direct assessment of knot performance is comparison of the correlation function of the parent process with that of the predictive process. Banerjee et al. (2008) report the functions to agree better at larger distances and even more so with increasing smoothness and range. What matters is the size of the range relative to the spacing of the grid for the knots. A practical rule is to keep the distance between adjacent knots smaller than the spatial range roughly gleaned from exploratory variograms. In principle, we recommend implementing the analysis over different choices of  $m$ ; we assess robustness of inference using several metrics to evaluate the stability of inference for varying knot densities, including deviance information criterion (DIC), parameters’ credible interval, and visual comparison between the resulting residual surfaces.

**3. Bayesian Implementation**

For the full model (2), we complete the Bayesian specification by assigning prior distributions to each parameter. Our data equation now resembles

$$\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{a} + \mathbf{d} + \mathbf{w} + \boldsymbol{\epsilon}, \tag{4}$$

where  $\boldsymbol{\epsilon}$ ,  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{w}$  are  $\text{MVN}(\mathbf{0}, \tau^2 I)$ ,  $\text{MVN}(\mathbf{0}, \sigma_a^2 \mathbf{A})$ ,  $\text{MVN}(\mathbf{0}, \sigma_d^2 \mathbf{D})$ , and  $\text{MVN}\{\mathbf{0}, \sigma_w^2 R(\boldsymbol{\theta})\}$ , respectively. Given priors, estimation employs a Gibbs sampler with a slice or metropolis steps. The inverse gamma (IG) prior, with shape  $\alpha$  and scale  $\eta$  parameters, is a conjugate choice for the variance parameters’ prior distribution and we set  $\sigma_a^2 \sim \text{IG}(\alpha_a, \eta_a)$ ;  $\sigma_d^2 \sim \text{IG}(\alpha_d, \eta_d)$ ;  $\sigma_w^2 \sim \text{IG}(\alpha_w, \eta_w)$ ; and  $\tau^2 \sim \text{IG}(\alpha_\epsilon, \eta_\epsilon)$ . Finally, we must assign priors to the parameters in  $\boldsymbol{\theta}$ : for the anisotropic Matérn function, priors are assigned to the spatial decay, smoothness, and anisotropic rotation parameter.

The spatial decay parameters are often weakly identifiable and reasonably informative priors are needed for satisfactory MCMC behavior. Priors for the decay parameters  $\lambda_1$  and  $\lambda_2$  are set relative to the size of the domain, e.g., setting the spatial ranges around a chosen fraction of the maximum distance. The smoothness parameter  $\nu$  is typically assigned a prior support in the interval  $(0, 2)$  (Stein, 1999), whereas the rotation parameter  $\psi$  can be assigned support in  $(-\pi/2, \pi/2)$  or  $[-\pi/2, \pi/2)$ .

In general, a sampler that works on the marginalized likelihood, integrating out the random and spatial effects, is likely to lead to quicker convergence. However, this would involve the  $n \times n$  matrix  $\sigma_a^2 \mathbf{A} + \sigma_d^2 \mathbf{D} + \sigma_w^2 R(\boldsymbol{\theta}) + \tau^2 I$  and does not offer computational gains. Alternatively, we design a sampler that block updates  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{w}$  from their full conditional distributions in each iteration of the Gibbs sampler. Note that the full conditionals for the  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{w}$  are multivariate normal distributions with dispersion matrices  $(\frac{1}{\sigma_a^2} \mathbf{A}^{-1} + \frac{1}{\tau^2} I)^{-1}$ ,  $(\frac{1}{\sigma_d^2} \mathbf{D}^{-1} + \frac{1}{\tau^2} I)^{-1}$ , and  $(\frac{1}{\sigma_w^2} R^{-1}(\boldsymbol{\theta}) + \frac{1}{\tau^2} I)^{-1}$ , respectively. In practice, we do not compute the inverse but some matrix decomposition that also yields determinants. The spectral decomposition is a natural choice: for  $\mathbf{a}$ , we have  $(\frac{1}{\sigma_a^2} \mathbf{A}^{-1} + \frac{1}{\tau^2} I)^{-1} = \mathbf{P}^T (\frac{1}{\sigma_a^2} \Lambda^{-1} + \frac{1}{\tau^2} I)^{-1} \mathbf{P}$ , where  $\mathbf{A} = \mathbf{P}^T \Lambda \mathbf{P}$  is the spectral decomposition of  $\mathbf{A}$ . This *trick* was also used by Thompson and Shaw (1990) for maximum likelihood estimation of variance components. Because the matrix  $\mathbf{A}$  is assumed known and fixed, this spectral decomposition is needed only once before the MCMC starts. Within the chains as  $\sigma_a^2$  and  $\tau^2$  are updated from their inverse-gamma full conditionals, one only needs to update  $n$  elements of a diagonal matrix. Exactly the same benefits apply to  $\mathbf{d}$ , but for  $\mathbf{w}$  the correlation matrix  $R(\boldsymbol{\theta})$  itself needs updating in each iteration and we resort to the methods of Section 2.1.

As mentioned in Section 2.1.1, the OU model  $R^{-1}(\phi_1, \phi_2)$  is a Kronecker product of tridiagonal matrices whose eigenvalues are computed in linear time; further gains are accrued by noting that  $R^{-1}(\phi_1, \phi_2) = (\mathbf{P}_1^T \otimes \mathbf{P}_2^T)(\Lambda_1^{-1} \otimes \Lambda_2^{-1})(\mathbf{P}_1 \otimes \mathbf{P}_2)$ , where  $R_i^{-1}(\phi_i) = \mathbf{P}_i^T \Lambda_i^{-1} \mathbf{P}_i$  for  $i = 1, 2$  (Martin, 1990). For the predictive process model and its modification, on the other hand, we replace  $\mathbf{w}$  in equation (4) by  $\tilde{\mathbf{w}}$  or  $\tilde{\mathbf{w}}_\epsilon$  (Section 2.1.2) requiring only  $m \times m$  dense matrix computations. See Web Appendix A for the explicit full conditional distributions and further details of the proposed algorithms.

In breeding trials phenotypic traits of the parent breeding stock and offspring are commonly unknown or missing. From a spatial perspective, this essentially means that we have an irregular arrangement of observed locations. The predictive process adapts smoothly to this scenario, treating the observed locations as simply irregularly arranged sites, but the OU method is precluded from direct implementation and we must sample from the predictive distribution of a missing response. These are easily generated as the complete conditional distribution is given by the model’s likelihood. For example, given the full model (2), the  $l$ th sample from the predictive distribution of the  $i$ th missing  $Y$  is easily computed as  $Y_i^{(l)} = X\boldsymbol{\beta}_i^{(l)} + a_i^{(l)} + d_i^{(l)} + w_i^{(l)} + \epsilon^{(l)}$  where  $\epsilon^{(l)}$  is a  $N(0, \tau^{2(l)})$  draw (see, e.g., Schafer, 1997). This will, however, significantly detract from the computational benefits highlighted in Section 2.1.1.

To compare several alternative models with varying degrees of genetic and spatial richness, we use the DIC (Spiegelhalter et al., 2002). Letting  $\Omega$  be the generic set of parameters being estimated for each model (including random effects), we compute the expected posterior deviance  $\overline{D(\Omega)} = E_{\Omega|\mathbf{Y}}\{-2\log L(\text{Data}|\Omega)\}$ , where  $L(\text{Data}|\Omega)$  is the first stage Gaussian likelihood from the respective model and the effective number of parameters (as a penalty) as  $p_D = \overline{D(\Omega)} - D(\bar{\Omega})$ , where  $\bar{\Omega}$  is the posterior mean of the model parameters. The DIC is then given by  $\overline{D(\Omega)} + p_D$  and is easily computed from the posterior samples with lower values indicating better models.

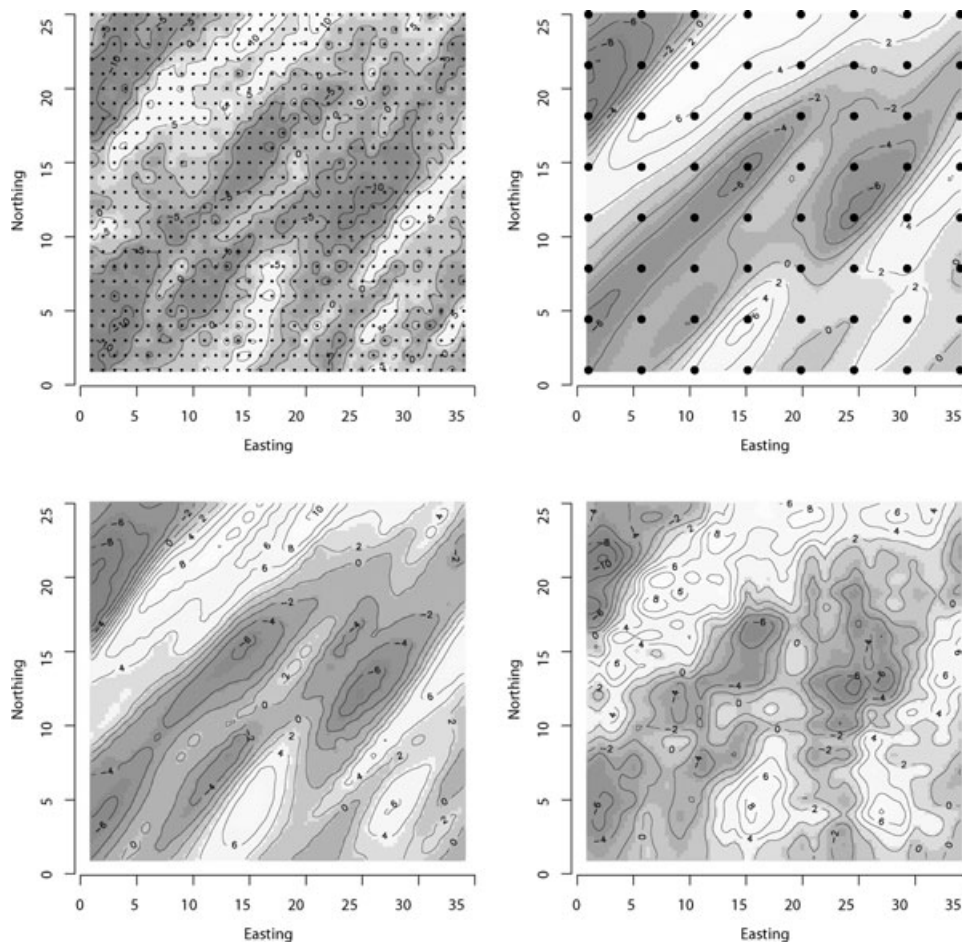
#### 4. Analysis of Data

##### 4.1 Synthetic Data Analysis

One replicate of full-sib families (800 offspring in total) were constructed by crossing 50 unrelated parents according to a partial diallele design (see Kempthorne and Curnow, 1961, for design details). The 850 individuals were assigned randomly to intersections of a  $34 \times 25$  unit grid. The synthetic population mean was set to 50 and the random polygenic effects

$\mathbf{a}$  and  $\mathbf{d}$  were drawn from  $\text{MVN}(\mathbf{0}, \sigma_a^2 \mathbf{A})$  and  $\text{MVN}(\mathbf{0}, \sigma_d^2 \mathbf{D})$ , with  $\sigma_a^2 = 25$  and  $\sigma_d^2 = 10$ . The random spatial effects,  $\mathbf{w}$ , were drawn from  $\text{MVN}\{\mathbf{0}, \sigma_w^2 R(\boldsymbol{\theta})\}$ , where  $\sigma_w^2 = 30$  with  $R(\boldsymbol{\theta})$  comprising anisotropic Matérn correlations with  $\nu = 0.5$ ,  $\lambda_1 = 14$ ,  $\lambda_2 = 3$ , and  $\psi = \pi/4$ . These yield an expected effective spatial range of 42, which is oriented to  $45^\circ$ , and a second spatial range of 9, perpendicular to the  $45^\circ$  axis. Here, effective range is defined as the distance at which the correlation drops to 0.05 (i.e., when  $\nu = 0.5$  the effective range is  $-\log(0.05)\lambda$ ). The measurement error,  $\boldsymbol{\epsilon}$ , was drawn from  $\text{MVN}(\mathbf{0}, \tau^2 I)$ , with  $\tau^2 = 10$ . The phenotypic values,  $\mathbf{Y}$ , for each of the 800 individuals were then generated by model (2). The top left image plot in Figure 1 illustrates the planting grid and the interpolated surface of the synthetic spatial random effects. The phenotypic values of the 50 unrelated parents are assumed unknown and are therefore treated as missing data in the subsequent analysis.

The synthetic dataset illustrates the extent to which the proposed samplers can recover genetic parameters of interest given nonspatial models, several predictive (and modified predictive) process models with isotropic and anisotropic



**Figure 1.** First row: Interpolated surface of random spatial effects associated with the 850 synthetic data points (left). Interpolated surface of the estimated random spatial effects based on 64 knots located at large points (right). Second row: Interpolated surface of estimated random effects for 256 knots (left) and OU (right) knots. This figure appears in color in the electronic version of this article.

**Table 1**  
*Model comparisons using the DIC criterion for  
the synthetic dataset*

Model	$pD$	DIC
<i>Nonspatial</i>		
Add.	274.53	4363.46
Add. Dom.	317.85	4333.35
<i>Spatial isotropic predictive process</i>		
64 knots	420.72	4036.80
144 knots	484.35	3573.45
256 knots	455.87	3564.03
<i>Spatial isotropic modified predictive process</i>		
64 knots	523.11	4014.34
144 knots	548.52	3562.07
256 knots	625.28	3513.65
<i>Spatial anisotropic predictive process</i>		
64 knots	431.48	3996.27
144 knots	503.07	3781.68
256 knots	526.13	3527.56
<i>Spatial anisotropic modified predictive process</i>		
64 knots	538.36	3807.72
144 knots	579.62	3547.60
256 knots	636.89	3490.12
<i>OU process model</i>	512.45	3156.03

covariance structures, and different knots and the separable OU processes. Assuming ignorance about the partitioning of the variance in  $\mathbf{Y}$ , we assign an  $IG(2, 20)$  prior to each variance parameter. The maximum distance across the synthetic domain is about 40; hence, for the anisotropic Matérn models, we assign  $\lambda_i \sim U(0.5, 20)$  for  $i = 1, 2$ , which corresponds to an effective range between 1.5 and 60 distance units (when  $\nu = 0.5$ ). The smoothness and rotation parameters,  $\nu$  and  $\psi$ , were assigned  $U(0, 2)$  and  $U[\pi/2, -\pi/2]$  priors, respectively. For the isotropic Matérn models, we assigned  $\phi \sim U(0.05, 2)$  and  $\nu \sim U(0, 2)$ , which together provide the same effective range support as the anisotropic models. For the separable OU processes, the row and column range parameters,  $\phi_r$  and  $\phi_c$  were also assigned  $U(0.05, 2)$  priors.

Using our proposed models and samplers we fit the 15 models listed in Table 1. For each model, three parallel MCMC chains were run for 25,000 iterations. The samplers were coded in C++ and leveraged threaded BLAS and LAPACK routines for matrix computations. The code was run on an Altix 350 SGI with two 64-bit 1.6 GHz Intel Itanium processors. Each chain of 25,000 iterations took less than 10 hours to complete. The CODA package in R ([www.r-project.org](http://www.r-project.org)) was used to diagnose convergence by monitoring mixing using Gelman–Rubin diagnostics and autocorrelations (see, e.g., Gelman et al., 2004, Section 11.6). Acceptable convergence was diagnosed within 10,000 iterations and therefore 45,000 samples (15,000  $\times$  3) were retained for posterior analysis.

Returning to Table 1, the DIC scores clearly indicate the importance of the dominance and spatial random effects. As expected, for both the isotropic and anisotropic predictive process models, increasing the number of knots used to model the parent process improved model fit. Despite the fact that the synthetic spatial process is strongly anisotropic, the DIC scores are only marginally lower than most of the correspond-

ing isotropic models. Comparing between the predictive process models, we see that despite the increased effective number of parameters,  $pD$ , resulting from the additional parameter vector  $\epsilon$ , the modified predictive process models have lower DIC scores than the unmodified predictive process models. Interestingly, the DIC score for the OU process model was appreciably lower than the anisotropic predictive process models with 256 knots—perhaps due to the OU process better adapting to local features than the smoother predictive process. Also, the anisotropic Matérn’s richer parameterization induces greater uncertainty in the predictive process models, whereas the OU process does induce anisotropy but with a simpler parameterization.

For predictive process models, increasing knot intensity allows the estimated random spatial effects to respond better to the residual spatial pattern. This trend is evident in the estimated random spatial effect surfaces for 64 and 256 knots in Figure 1. The 64-knot overlay is illustrated in the top right image plot in Figure 1; however, to reduce clutter, the 256 knot locations are not added to the lower left image plot. Figure 1 shows (from the contour orientation) that the predictive process captures the true anisotropy and spatial ranges much more effectively than the OU process. Visually, there is negligible difference between the spatial effect surfaces and therefore, only the modified predictive processes are illustrated in Figure 1. This is expected as the former derives directly from the underlying anisotropic Matérn family, whereas the latter does not. Still, the OU process does reveal some of the underlying anisotropy (contours displaying some 45° orientation) and, as mentioned above, yields a better goodness of fit.

Web Table 1 offers parameter estimates of the two nonspatial models. The variance associated with missing random effects (e.g., dominance and/or spatial) is “captured” by  $\tau^2$ . Specifically, for the additive effects model, the estimated posterior median of  $\tau^2$  is 45.57 with lower and upper 95% credible intervals of 34.53 and 54.60, respectively. For the additive and dominance effects model, the estimated posterior median of  $\tau^2$  is 42.2 with lower and upper 95% credible intervals of 28.12 and 51.92, respectively. Neither interval covers the “true” synthetic value of 10. The consequence, besides violating the basic assumption of uncorrelated residuals, is that narrow-sense heritability,  $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_d^2 + \tau^2)$ , and proportion of dominance,  $d^2 = \sigma_d^2 / (\sigma_a^2 + \sigma_d^2 + \tau^2)$ , are downwardly biased and the individuals’ breeding values could be incorrect. Despite the above violation in the nonspatial models, the  $\sigma_a^2$  and  $\sigma_d^2$  are remarkably close to the “true” synthetic values of 25 and 10, respectively. Table 2 provides parameter estimates for the predictive process models. Here, the credible intervals associated with the spatial effect parameters (i.e.,  $\sigma_w^2$ ,  $\nu$ ,  $\psi$ ,  $\lambda_1$ , and  $\lambda_2$ ) include the true values for all knot intensities. Expectedly, increasing knot intensity beyond 64 knots tightens these credible intervals. Importantly, the upward bias induced by the unmodified predictive process is evident when comparing the  $\tau^2$  estimates to those from the modified process. Further, we see that this bias is inversely related to the knot intensity, i.e., the lower the knot intensity the higher the  $\tau^2$  bias. Web Table 2 shows the parameter estimates for the OU model. Despite its limitations, the OU model’s estimated credible intervals for the intercept and variance parameters include the

**Table 2**

Parameter credible intervals, 50% (2.5%, 97.5%) for the predictive process and modified predictive process anisotropic models each with three knot intensities across the synthetic domain. Credible intervals that do not include the parameter's true value are bolded.

Parameter	True	Predicted process			Modified predicted process		
		64 knots	144 knots	256 knots	64 knots	144 knots	256 knots
$\beta$	50	49.86 (46.68, 53.37)	49.50 (46.07, 52.72)	49.37 (45.85, 52.41)	49.60 (46.79, 52.86)	49.32 (45.31, 52.43)	50.02 (46.38, 52.53)
$\sigma_a^2$	25	27.17 (17.46, 42.10)	30.25 (19.02, 46.72)	32.20 (19.52, 48.85)	27.06 (17.59, 41.93)	29.01 (18.49, 44.80)	30.64 (19.70, 46.51)
$\sigma_d^2$	10	4.02 (1.43, 11.45)	5.08 (1.65, 13.40)	7.10 (2.05, 18.58)	4.29 (1.45, 12.98)	4.72 (1.72, 12.68)	4.39 (1.52, 13.70)
$\sigma_g^2$	30	27.16 (15.37, 51.13)	30.79 (18.73, 51.10)	29.87 (18.58, 47.92)	26.48 (14.90, 39.19)	30.62 (19.22, 53.40)	33.96 (20.33, 54.49)
$\tau^2$	10	<b>24.65 (15.92, 33.23)</b>	17.87 (8.75, 26.53)	13.35 (4.28, 24.73)	16.73 (8.67, 26.62)	12.74 (5.83, 21.68)	10.33 (4.01, 19.62)
$\nu$	0.5	0.48 (0.23, 0.96)	0.64 (0.32, 0.86)	0.65 (0.35, 0.87)	0.76 (0.33, 0.93)	0.69 (0.32, 0.89)	0.59 (0.31, 0.85)
$\psi$	45°	38.43 (27.46, 51.56)	39.76 (32.00, 48.22)	41.44 (34.36, 48.72)	38.03 (27.72, 49.98)	39.44 (32.03, 48.12)	40.79 (23.08, 57.82)
$\lambda_1$	14	15.00 (8.05, 17.88)	13.43 (7.44, 17.78)	13.50 (7.54, 17.75)	15.13 (8.37, 17.90)	14.35 (7.43, 17.84)	12.27 (4.98, 17.74)
$\lambda_2$	3	3.48 (1.31, 6.47)	2.95 (1.55, 5.14)	3.01 (1.73, 4.99)	3.32 (1.16, 6.35)	3.08 (1.59, 5.57)	3.28 (1.53, 5.21)
Eff. range <sub>1</sub>	42	50.66 (25.61, 68.08)	43.27 (23.11, 65.02)	44.16 (23.01, 64.26)	51.29 (26.03, 68.67)	47.04 (23.04, 67.18)	38.89 (14.16, 63.89)
Eff. range <sub>2</sub>	9	11.89 (4.23, 22.20)	9.69 (4.72, 16.89)	9.89 (5.28, 16.34)	11.54 (3.75, 22.54)	10.36 (4.74, 19.51)	10.57 (4.53, 18.57)
$h^2$	0.55	0.48 (0.33, 0.66)	0.56 (0.38, 0.75)	0.60 (0.40, 0.79)	0.54 (0.37, 0.72)	0.61 (0.43, 0.80)	0.66 (0.47, 0.83)
$d^2$	0.22	<b>0.07 (0.02, 0.20)</b>	0.09 (0.03, 0.25)	0.13 (0.04, 0.35)	0.08 (0.03, 0.26)	0.10 (0.04, 0.27)	0.09 (0.03, 0.29)

true values. Importantly, by including random spatial effects through the OU process or modified predictive process, the credible intervals for  $\tau^2$  include the “true” value of 10 from which the data was simulated; hence,  $h^2$  more accurately reflects the potential success of breeding efforts.

Ultimately, interest turns to inference on individuals’ breeding value and phenotype prediction. For the entire set of 800 offspring, the OU and predictive process models provided credible intervals that contained the true breeding value about 95% of the time for the genetic and spatial effect vectors. Web Figure 1 shows posterior credible intervals for a subset of the true additive, dominance, and spatial effects estimated by the OU model.

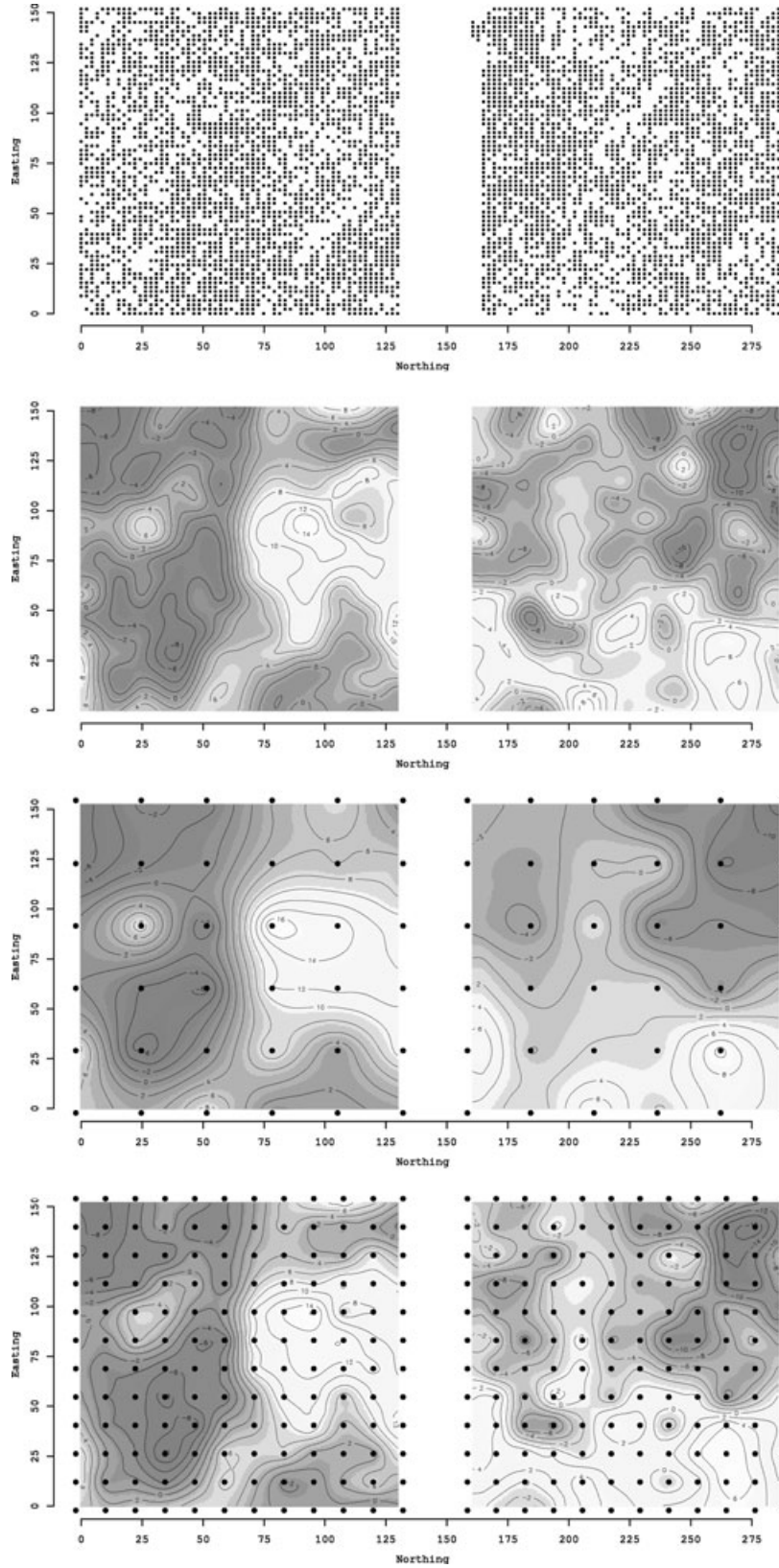
4.2 Scots Pine Data Analysis

Field measurements from a 26-year-old Scots pine (*Pinus sylvestris* L.) progeny study in northern Sweden serves as a trial dataset for our proposed methods. The original design established in 1971 (by Skogfors; trial identification S23F7110264 Vindeln) consisted of 8160 seedlings divided into two spatially disjoint trial sites (i.e., denoted as northern and southern trial site), see Figure 2. Trial sites were subdivided into  $8.8 \times 22$  m blocks each containing 40 seedlings placed on a  $2.2 \times 2.2$  m grid. The northern site held 105 blocks and the southern site had 99 blocks. Parent stock was crossed according to a partial diallel design of 52 parent trees assumed unrelated and seedlings were planted out on the grids unrestricted randomly. In 1997 the 4970 surviving trees were measured for various traits of breeding interest, including total tree height. For a description of the full dataset, see Waldmann and Ericsson (2006).

Two analyses were conducted with these data. First, using only the northern trial site we compare genetic nonspatial, OU, and predictive process models. Next, we pool the northern and southern trial sites and fit the genetic nonspatial and predictive process models. The spatial discontinuity of the two trial sites precludes application of the OU. We present the analysis and results of the northern trial site as a supplementary analysis in Web Appendix B, whereas those for the full dataset are detailed in the subsequent sections.

We fit eight candidate models of varying complexity to the Scots pine data. The first assumes that the variance in the observed trait can be explained by only an intercept and additive random effect. Building on the first model, the second includes the dominance random effect. In addition to the genetic random effects, we considered the isotropic unmodified and modified predictive process models, each with three knot intensities, 72, 128, and 288. An initial analysis, using directional semivariograms (see Web Figure 2) and fitting the modified predictive anisotropic model with 288 knots, revealed no directional pattern of dependence in the residuals of the additive and dominance random effect model. This lack of residual directional spatial dependence agrees with the irregular soil structures on such moraine sites in Sweden. Therefore, the remainder of the analysis focuses on only the isotropic models.

To complete the Bayesian specification of the candidate models, we assign priors and hyperpriors to the parameters. A semivariogram of observed tree height aids in hyperprior specifications (see Web Figure 2). The nugget of the semivariogram suggests that the sum of the genetic and random



**Figure 2.** First row: Each point represents a Scots pine tree location. Second row: Interpolated surface of model (1) residuals for stem height. Third and fourth rows: Estimated surface of spatial random effects with 72 and 288 knot intensities (large points) modeled with the stationary isotropic Matérn correlation function. This figure appears in color in the electronic version of this article.

**Table 3**  
Model comparisons using the DIC criterion for the Scots pine dataset

Model	$p_D$	DIC
<i>Nonspatial</i>		
Add.	648.37	30,322.03
Add. dom.	1017.35	30,209.83
<i>Spatial isotropic predictive process</i>		
72 knots	1456.54	28,995.19
128 knots	1410.17	28,945.41
288 knots	1368.74	28,906.74
<i>Spatial isotropic modified predictive process</i>		
72 knots	1924.39	28,763.88
128 knots	2053.99	28,569.56
288 knots	1981.42	28,563.44

variance is  $\sim 130$  and a spatial variance (i.e., partial-sill) is  $\sim 50$ . Thus, in the full model (2), we assign  $IG(2, 40)$  priors to  $\sigma_a^2$ ,  $\sigma_d^2$ , and  $\tau^2$ , and an  $IG(2, 50)$  to  $\sigma_w^2$ . Further, the semivariogram suggests an effective range of about  $\sim 110$  meters, hence, for the isotropic Matérn, we assigned  $\phi \sim U(0.02, 0.14)$ . We reduce the computational demand by fixing  $\nu$  at 0.5. Therefore, given the prior on  $\phi$ , the models support an effective spatial range of about 20 to 150 meters. We assume a flat prior for the intercept. For each candidate model, we ran three parallel MCMC chains for 25,000 iterations, which took less than 24 hours. Again, visual inspection and statistical diagnostics suggested acceptable convergence within 10,000 iterations, and 45,000 samples ( $15,000 \times 3$ ) were retained for posterior analysis.

DIC scores, Table 3, and the nonspatial models parameter estimates, Web Table 3, support adding the dominance effects to the simple additive model. Next, DIC scores and the nonspatial model residual surface plot, Figure 2, and associated semivariogram, Web Figure 2, exhibit clear spatial dependence and thus support the addition of random spatial effects. The decreasing DIC scores in Table 3 show that the predictive process models improve model fit over the nonspatial models. Not surprisingly, the DIC scores for both predictive process models support increasing knot intensity. Further, as in the synthetic data analysis, DIC scores favor the modified predictive process despite the increased number of effective parameters. Next, Table 4 provides the parameter estimates for the predictive process models. Comparing these parameter estimates to the nonspatial model estimates, we again see that the addition of random spatial effects decreases  $\tau^2$  and subsequently increases  $h^2$ . Reductions in computational demand afforded by the predictive process models come at a price. Figure 2 illustrates the result of reducing knot intensity from 288 (fourth row) to 72 (third row), i.e., causing a smoothing of the estimated spatial process. Importantly, however, spatial correlation function parameters and variance estimates do not change much as the number of knots decreases. Finally, as in the synthetic data analysis, we see the bias reduction provided by the predictive process modification and note that the bias increases with decreasing knot intensity for the unmodified predictive process models.

**Table 4**  
Parameter credible intervals, 50% (2.5%, 97.5%) for the predictive process and modified predictive process isotropic models each with three knot intensities across the Scots pine trial

Parameter	Predicted process			Modified predictive process		
	72 knots	128 knots	288 knots	72 knots	128 knots	288 knots
$\beta$	69.83 (66.34, 73.98)	70.16 (65.38, 73.35)	69.39 (65.99, 74.16)	68.68 (64.66, 72.49)	69.35 (66.20, 73.17)	70.32 (66.39, 73.75)
$\sigma_a^2$	32.03 (21.28, 54.33)	33.26 (21.37, 54.17)	34.56 (23.51, 55.27)	32.80 (21.09, 54.47)	35.61 (21.25, 52.62)	35.60 (22.23, 59.18)
$\sigma_d^2$	16.91 (3.74, 34.82)	18.07 (9.92, 33.05)	17.42 (10.44, 29.40)	18.37 (10.55, 28.59)	16.38 (8.87, 32.94)	17.99 (5.65, 31.36)
$\sigma_w^2$	52.48 (35.72, 80.60)	50.59 (35.18, 74.42)	47.20 (34.57, 67.98)	45.47 (32.81, 64.33)	52.60 (31.99, 77.38)	50.14 (36.84, 70.12)
$\tau^2$	99.73 (85.43, 113.77)	95.60 (82.38, 105.86)	92.97 (79.67, 103.87)	79.18 (65.47, 91.82)	78.11 (63.43, 90.00)	80.52 (62.81, 92.41)
$\phi$	0.04 (0.02, 0.06)	0.04 (0.02, 0.06)	0.04 (0.03, 0.06)	0.04 (0.02, 0.06)	0.04 (0.02, 0.06)	0.04 (0.02, 0.07)
Eff. range	85.03 (52.01, 138.24)	74.52 (48.98, 125.71)	69.20 (47.00, 117.66)	81.27 (52.37, 136.73)	74.76 (48.05, 130.93)	69.39 (45.43, 121.53)
$h^2$	0.21 (0.15, 0.34)	0.22 (0.15, 0.34)	0.24 (0.17, 0.35)	0.25 (0.16, 0.39)	0.27 (0.17, 0.38)	0.27 (0.17, 0.41)
$d^2$	0.11 (0.02, 0.23)	0.12 (0.07, 0.22)	0.12 (0.07, 0.20)	0.14 (0.08, 0.22)	0.13 (0.07, 0.25)	0.13 (0.04, 0.23)

## 5. Discussion

This article offers computationally efficient methods for Bayesian estimation of genetic variance components in the individual-based linear mixed model with or without inclusion of spatial random effects. Both the OU and modified predictive process models provided comparable genetic and spatial parameter estimates in the synthetic and Scots pine northern trial site analyses. The synthetic analysis showed that the OU model provides robust estimates even when the underlying spatial process is characterized by strong anisotropy, misaligned with the grid's rows and columns. However, requiring observations on an uninterrupted regular grid severely limits the use of OU models for many problems encountered in practice. In Section 3, we offered an approach to deal with missing data, but the pervasive limitation of gridded observations remains, which prevented the OU process models from being used for analysis of the full Scots pine dataset (i.e., pooled northern and southern trial sites). In addition, a computationally effective OU model has inherent modeling limitations with sparse analytical inverses of  $R(\theta)$  being available only for special cases. This precludes richer correlations functions that may better describe association among proximate observations (e.g., nonstationary anisotropic Matérn correlation). Unlike the OU model, the modified predictive process accommodates complex spatial correlation structures and nongridded trial designs. With 4970 observations it is computationally challenging to fit a traditional Bayesian hierarchical geostatistical model to the full Scots pine dataset. The predictive process approach provides one viable alternative that does not require analysts making unreasonable assumptions or sacrificing model complexity. It is especially suitable for random effects models, such as here, where the process is completely unobserved and empirical estimates (as needed by many existing approaches, e.g., Cressie and Johannesson, 2008) are difficult to obtain. Further, all illustrations presented here demonstrate that the modified predictive process substantially reduced the upward bias of  $\tau^2$  that accompanies the predictive process. Effectively negating this bias is critical when interest lies in partitioning variance and calculating derived metrics such as  $h^2$  and  $d^2$ .

## 6. Supplementary Materials

Web Appendices, Figures, and Tables referenced in Sections 1, 3, 4.1, and 4.2 are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

## ACKNOWLEDGEMENTS

The authors thank the referees, the associate editor, and the editor for their valuable comments and suggestions. The authors also thank Dr Huiyan Sang for several useful discussions regarding the predictive process. The first two authors would also like to acknowledge the support of the National Science Foundation's DMS-0706870 grant.

## REFERENCES

- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman and Hall/CRC Press.

- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society, Series B* **70**, 825–848.
- Cappa, E. P. and Cantet, R. J. C. (2007). Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. *Canadian Journal of Forest Research* **37**, 2677–2688.
- Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variances for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882.
- Cressie, N. (1993). *Statistics for Spatial Data*, 2nd edition. New York: Wiley.
- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B* **70**, 209–226.
- Cullis, B. R., Gogel, B., Verbyla, A., and Thompson, R. (1998). Spatial analysis of multi-environment early generation variety trials. *Biometrics* **54**, 1–18.
- Dutkowski, G. W., Silva, J. C., Gilmour, A. R., and Lopez, G. A. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* **32**, 2201–2214.
- Fuentes, M. (2002). Periodogram and other spectral methods for nonstationary spatial processes. *Biometrika* **89**, 197–210.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* **102**, 321–331.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd edition. Boca Raton, Florida: Chapman and Hall/CRC Press.
- Henderson, C. R. (1985). MIVQUE and REML estimation of additive and nonadditive genetic variances. *Journal of Animal Science* **61**, 113–121.
- Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*, C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi (eds), 37–56. London: Springer-Verlag.
- Kamman, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics* **52**, 1–18.
- Kempthorne, O. and Curnow, R. N. (1961). The partial diallel cross. *Biometrics* **17**, 229–250.
- Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, Massachusetts: Sinauer Associates.
- Martin, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics—Theory and Methods* **19**, 55–81.
- Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large datasets. *Computational Statistics and Data Analysis* **51**, 3631–3653.
- Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**, 483–506.
- Rue, H. and Held, L. (2006). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, Florida: Chapman and Hall/CRC Press.
- Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics* **29**, 31–49.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Boca Raton, Florida: Chapman and Hall/CRC Press.
- Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. New York: Springer-Verlag.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and 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.

- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory of Kriging*. New York: Springer.
- Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating likelihoods for large spatial datasets. *Journal of the Royal Statistical Society, Series B* **66**, 275–296.
- Thompson, E. A. and Shaw, R. G. (1990). Pedigree analysis for quantitative traits: Variance components without matrix inversion. *Biometrics* **46**, 399–413.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society, Series B* **50**, 297–312.
- Ver Hoef, J. M., Cressie, N. A. C., and Barry, R. P. (2004). Flexible spatial models based on the fast Fourier transform (FFT) for cokriging. *Journal of Computational and Graphical Statistics* **13**, 265–282.
- Waldmann, P. and Ericsson, T. (2006). Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theoretical and Applied Genetics* **112**, 1441–1451.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–439.
- Zimmerman, D. L. and Harville, D. A. (1991). A random field approach to the analysis of field–plot experiments and other spatial experiments. *Biometrics* **47**, 223–239.

*Received June 2007. Revised May 2008.*

*Accepted May 2008.*