Nonstationary Spatial Gaussian Markov Random Fields

Yu YUE and Paul L. SPECKMAN

Thin-plate splines have been widely used as spatial smoothers. In this article, we present a Bayesian adaptive thin-plate spline (BATS) suitable for modeling nonstationary spatial data. Fully Bayesian inference can be carried out through efficient Markov chain Monte Carlo simulation. Performance is demonstrated with simulation studies and by an application to a rainfall dataset. A FORTRAN program implementing the method, a proof of the theorem, and the dataset in this article are available online.

Key Words: Block sampling; Intrinsic autoregressive; Sparse structure; Spatially adaptive.

1. INTRODUCTION

Spatial smoothing has been one focus of spatial statistics research. Statisticians have been interested in estimating a smooth spatial process from noisy observations, constructing smoothed maps, and predicting at locations for which data are missing. Two prominent approaches are thin-plate splines and kriging (see Cressie 1993, for a review), and the connection between those two was nicely developed by Nychka (2000). The thin-plate spline is a multidimensional extension of the smoothing spline (e.g., Wahba 1990; Green and Silverman 1994; Gu 2002), and its two-dimensional version has been widely used as a spatial smoother.

Consider a two-component nonparametric regression model

$$y_i = f(u_i, v_i) + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{1.1}$$

where f is an unknown but smooth function and ε_i is a mean-zero noise term. The thin-plate spline estimator of f in (1.1) is the solution to the minimization problem

$$\hat{f} = \arg\min_f \left[\sum_{i=1}^{N} (y_i - f(u_i, v_i))^2 + \lambda J_2(f) \right], \tag{1.2}$$

Yu Yue is Assistant Professor, Department of Statistics and CIS, Baruch College, City University of New York, New York, NY 10010. Paul L. Speckman is Professor, Department of Statistics, University of Missouri, Columbia, MO 65211 (E-mail: *speckmanp@missouri.edu*).

Journal of Computational and Graphical Statistics, Volume 19, Number 1, Pages 96–116
DOI: 10.1198/jcgs.2009.08124

where $J_2(f)$ is the roughness penalty for function f on $\mathbb{R}^2$,

$$J_2(f) = \int\int_{\mathbb{R}^2} \left[\left(\frac{\partial^2 f(u,v)}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 f(u,v)}{\partial u \, \partial v} \right)^2 + \left(\frac{\partial^2 f(u,v)}{\partial v^2} \right)^2 \right] du \, dv, \qquad (1.3)$$

and the smoothing parameter λ controls the trade-off between fidelity to the data with respect to sum squared error and smoothness from the penalty term in (1.3). A global λ, however, fails to adapt to variable smoothness in the function of interest. This is of particular importance in spatial datasets, in which a nonstationary process is often detected.

To improve the spatial adaptivity of (1.2), we propose a prior extending a Bayesian version of discretized thin-plate splines, a type of intrinsic Gaussian Markov random field (GMRF) (e.g., Rue and Held 2005), by introducing a spatially adaptive variance component and taking a further GMRF prior for this variance function. Fully Bayesian inference, including estimation of function and local variance parameters, can be carried out through efficient Markov chain Monte Carlo (MCMC) techniques (Gelfand and Smith 1990). The approach proposed in the article is a two-dimensional extension of the adaptive modeling for the one-dimensional scenario in the works of Yue, Speckman, and Sun (2008) and Lang, Fronk, and Fahrmeir (2002).

A related alternative Bayesian model has been proposed by Brezger, Fahrmeir, and Hennerfeind (2007). They introduced a class of adaptive GMRFs with stochastic interaction weights in a space-varying coefficient model. The approach extended the intrinsic GMRFs by taking independent gamma priors on the unknown weights, which, however, failed to provide a spatially varying covariance as needed when estimating functions with high spatial variability (see, e.g., Lang, Fronk, and Fahrmeir 2002). One remedy is to let the logarithms of the weights follow another GMRF model in a second hierarchy, but now on the graph of links (edges) between regions (nodes), not on the original graph of regions. Knorr-Held (2003) and Brezger, Fahrmeir, and Hennerfeind (2007) discussed the potential issues with such a formulation due to the increased complexity of the model. First, the new graphs of links may become rather large whereas information in the data is typically limited; second, the normalizing constant of the adaptive GMRF, which is a function of the weights, is quite computationally intensive; third, propriety of the posterior distribution must be shown because the problem already arises in the nonadaptive case with an improper prior for variance components as discussed by Speckman and Sun (2003) and Yue, Speckman, and Sun (2008). The nonstationary GMRF proposed here overcomes those problems by constructing a particular GMRF with a full-rank factorization property in the first hierarchy and then taking another simpler GMRF prior for adaptive variance function.

Beginning with Ruppert and Carroll (2000), a series of spatially adaptive penalized spline models have been proposed by Lang and Brezger (2004), Baladandayuthapani, Mallick, and Carroll (2005), Crainiceanu et al. (2007), and Krivobokova, Crainiceanu, and Kauermann (2008). Those methods achieve spatial adaptivity by imposing a functional structure on the smoothing parameters in ordinary penalized splines. In work similar to ours, Lang and Brezger (2004) used tensor products of B-spline bases and Crainiceanu et al. (2007) used low-rank bivariate thin-plate spline bases to do Bayesian adaptive bivariate smoothing; Krivobokova, Crainiceanu, and Kauermann (2008) did a frequentist analy-

sis of those adaptive P-spline models using the Laplace approximation of the marginal likelihood for estimation.

Several other Bayesian spatially adaptive models based on splines have been proposed. Mackay and Takeuchi (1998), Di Matteo, Genovese, and Kass (2001), Holmes and Mallick (2001), and Wood, Jiang, and Tanner (2002) treated additive models with one-dimensional components. There is also a growing literature on implementing nonstationary covariances or spatially adaptive models in the kriging framework. This includes the work of Higdon, Swall, and Kern (1999), Fuentes and Smith (2001), Sampson and Guttorp (1992), Damian, Sampson, and Guttorp (2001), Schmidt and O'Hagan (2003), Paciorek and Schervish (2004, 2006), and Zhu and Wu (2010).

The rest of the article is organized as follows. Nonstationary GMRF priors with varying variances and the corresponding adaptive thin-plate splines are introduced in Section 2. Inference is fully Bayesian and uses MCMC techniques, combining Gibbs sampling with efficient block moves for estimating the response function and Metropolis–Hastings algorithms with conditional prior proposals for estimating the variance function. Details are given in Section 3. Performance is investigated by the simulation studies in Section 4, and an application to a U.S. precipitation dataset is given in Section 5.

2. BAYESIAN ADAPTIVE THIN-PLATE SPLINES ON A LATTICE

2.1 THIN-PLATE SPLINES ON A LATTICE

To start, we modify (1.1) by taking data on a regular lattice $(\tilde{u}_j, \tilde{v}_k)$,

$$y_{jk\ell} = f(\tilde{u}_j, \tilde{v}_k) + \varepsilon_{jk\ell}, \qquad \varepsilon_{jk\ell} \overset{\text{iid}}{\sim} N(0, \tau^{-1}),$$
$$j = 1, \ldots, n_1, k = 1, \ldots, n_2, \ell = 1, \ldots, r_{jk}, \tag{2.1}$$

where $\tilde{u}_j$ and $\tilde{v}_k$ are equally spaced in vertical and horizontal directions with distance h, n_1 and n_2 are the number of rows and columns, respectively, for the lattice, and r_{jk} is the number of observations at grid point $(\tilde{u}_j, \tilde{v}_k)$. The total number of observations is $N = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} r_{jk}$. Note that $r_{jk} = 0$ is possible, that is, observations are not required at every grid point. Now vectorize the model by taking $z_{jk} = f(\tilde{u}_j, \tilde{v}_k)$, $\mathbf{z} = \text{Vec}([z_{jk}]) = (z_{11}, \ldots, z_{n_1 1}, z_{12}, \ldots, z_{n_1 n_2})'$, $\mathbf{y} = (y_{111}, \ldots, y_{11r_{11}}, y_{211}, \ldots, y_{n_1 n_2 r_{n_1 n_2}})'$, and $\boldsymbol{\varepsilon}$ defined similarly to $\mathbf{y}$. Finally, let $\mathbf{D} = [d_{ij}]$ be the $N \times n$ ($n = n_1 n_2$) incidence matrix defined by $d_{ij} = 1$ if observation i is at location j in the vectorized ordering of the lattice defined in (2.1), and $d_{ij} = 0$ if observation i is at location other than j. For instance, $d_{12} = 1$ indicates that the first observation is at grid point $(\tilde{u}_2, \tilde{v}_1)$. Then we can vectorize (2.1) as

$$\mathbf{y} = \mathbf{D}\mathbf{z} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \text{N}_N(\mathbf{0}, \tau^{-1}\mathbf{I}_N), \tag{2.2}$$

where $\text{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The setup outlined so far is best suited to regularly spaced data. For nonregular data, we propose binning the data to a regular lattice. Binning has been proposed by a number of authors in nonparametric regression and density estimation (e.g., Härdle and Scott 1992; Scott 1992) and is particularly useful when dealing with large datasets.

2.2 AN INTRINSIC GMRF IN TWO DIMENSIONS

Bayesian inference of model (2.2) requires a prior on $\mathbf{z}$. We here construct a particular bivariate intrinsic GMRF prior based on an approximation of the *Laplacian* differential operator with additional boundary assumptions.

Let $C_0^4(\mathbb{R}^2)$ denote the space of functions on $\mathbb{R}^2$ with integrable derivatives of order up to 3, all of which vanish at infinity. For $f \in C_0^4(\mathbb{R}^2)$, we integrate by parts with the boundary conditions to obtain

$$\begin{aligned}
&\iint_{\mathbb{R}^2} \left(\frac{\partial^2 f(u,v)}{\partial u\, \partial v} \right)^2 du\, dv \\
&= \iint_{\mathbb{R}^2} \frac{\partial^2 f(u,v)}{\partial u\, \partial v} \frac{\partial^2 f(u,v)}{\partial u\, \partial v}\, du\, dv \\
&= \int \frac{\partial^2 f(u,v)}{\partial u\, \partial v} \frac{\partial f(u,v)}{\partial u}\bigg|_{-\infty}^{\infty} du - \iint_{\mathbb{R}^2} \frac{\partial^3 f(u,v)}{\partial u\, \partial v^2} \frac{\partial f(u,v)}{\partial u}\, du\, dv \\
&= -\int \frac{\partial^2 f(u,v)}{\partial v^2} \frac{\partial f(u,v)}{\partial u}\bigg|_{-\infty}^{\infty} dv + \iint_{\mathbb{R}^2} \frac{\partial^2 f(u,v)}{\partial u^2} \frac{\partial^2 f(u,v)}{\partial v^2}\, du\, dv \\
&= \iint_{\mathbb{R}^2} \frac{\partial^2 f(u,v)}{\partial u^2} \frac{\partial^2 f(u,v)}{\partial v^2}\, du\, dv. \qquad (2.3)
\end{aligned}$$

Substituting (2.3) into (1.3), the penalty term $J_2(f)$ becomes

$$\begin{aligned}
J_2(f) &= \iint_{\mathbb{R}^2} \left[\left(\frac{\partial^2 f(u,v)}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 f(u,v)}{\partial u^2} \right) \left(\frac{\partial^2 f(u,v)}{\partial v^2} \right) + \left(\frac{\partial^2 f(u,v)}{\partial v^2} \right)^2 \right] du\, dv \\
&= \iint_{\mathbb{R}^2} \left[\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2} \right) f(u,v) \right]^2 du\, dv. \qquad (2.4)
\end{aligned}$$

If we define a bivariate Laplacian differential operator $\Delta = (\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2})$, the solution to the minimization problem

$$\arg\min_f \left[\sum_{i=1}^{N} (y_i - f(u_i, v_i))^2 + \lambda \iint (\Delta f(u,v))^2\, du\, dv \right] \qquad (2.5)$$

is also a thin-plate spline estimator (see, e.g., Bookstein 1989).

The GMRF prior introduced here is motivated by a difference approximation of the penalty term in (2.5). Assuming h is small, the second partial derivatives of $f(\tilde{u}, \tilde{v})$ in model (2.1) at point $(\tilde{u}_j, \tilde{v}_k)$ can be approximated by

$$\frac{\partial^2}{\partial \tilde{u}^2} f(\tilde{u}_j, \tilde{v}_k) \approx h^{-2} \nabla^2_{(1,0)} f(\tilde{u}_j, \tilde{v}_k) \qquad \text{and} \qquad \frac{\partial^2}{\partial \tilde{v}^2} f(\tilde{u}_j, \tilde{v}_k) \approx h^{-2} \nabla^2_{(0,1)} f(\tilde{u}_j, \tilde{v}_k),$$

where $\nabla^2_{(1,0)}$ and $\nabla^2_{(0,1)}$ denote the second-order backward difference operators

$$\begin{aligned}
\nabla^2_{(1,0)} f(\tilde{u}_j, \tilde{v}_k) &= f(\tilde{u}_j, \tilde{v}_k) - 2 f(\tilde{u}_{j-1}, \tilde{v}_k) + f(\tilde{u}_{j-2}, \tilde{v}_k), \\
\nabla^2_{(0,1)} f(\tilde{u}_j, \tilde{v}_k) &= f(\tilde{u}_j, \tilde{v}_k) - 2 f(\tilde{u}_j, \tilde{v}_{k-1}) + f(\tilde{u}_j, \tilde{v}_{k-2}).
\end{aligned} \qquad (2.6)$$

Following the notation in (2.6), define a discrete approximation to the Laplacian $\Delta f(\tilde{u}_j, \tilde{v}_k)$ in (2.5) for model (2.1) by

$$h^{-2}\nabla_0^2 f(\tilde{u}_j, \tilde{v}_k) \triangleq h^{-2}\left(\nabla_{(1,0)}^2 + \nabla_{(0,1)}^2\right) f(\tilde{u}_{j+1}, \tilde{v}_{k+1}), \tag{2.7}$$

which can be regarded as an extension of the difference operator defined for univariate random walks (Patra and Karttunen 2006). Therefore, one straightforward approximation of the penalty term in (2.5) is

$$\frac{1}{h^4}\sum_{j=2}^{n_1-1}\sum_{k=2}^{n_2-1}\{\nabla_0^2 f(\tilde{u}_j, \tilde{v}_k)\}^2. \tag{2.8}$$

Recalling the notation $z_{jk} = f(\tilde{u}_j, \tilde{v}_k)$, $1 \le j \le n_1$, $1 \le k \le n_2$, the sum of squares in (2.8) can be written as $\mathbf{z}'\mathbf{A}_0\mathbf{z}$, where $\mathbf{A}_0$ is a particular semidefinite matrix whose entries are the coefficients in (2.8). If we let $\lambda_h = \lambda/h^4$, the vector $\hat{\mathbf{z}}$ defined by

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}}[(\mathbf{y} - \mathbf{D}\mathbf{z})'(\mathbf{y} - \mathbf{D}\mathbf{z}) + \lambda_h \mathbf{z}'\mathbf{A}_0\mathbf{z}] \tag{2.9}$$

is a discretized thin-plate spline estimator. The optimization criterion (2.9) suggests the prior on $\mathbf{z}$ for Bayesian modeling of thin-plate splines should be of the form

$$[\mathbf{z}|\delta] \propto \delta^{(n-m)/2}|\mathbf{A}_0|_+^{1/2}\exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}_0\mathbf{z}\right), \tag{2.10}$$

where m is the dimension of the null space of $\mathbf{A}_0$ and δ is a precision that needs to be specified or estimated. The determinant $|\mathbf{A}_0|_+$ is the product of the nonzero eigenvalues of $\mathbf{A}_0$ and is irrelevant in Bayesian computation. Moreover, $\mathbf{A}_0$ is a so-called "structure matrix" that can be factored as $\mathbf{A}_0 = \mathbf{B}_0'\mathbf{B}_0$, where $\mathbf{B}_0$ is an $(n_1 - 2)(n_2 - 2) \times n$ full-rank matrix with the coefficients of $\nabla_0^2 f(\tilde{u}, \tilde{v})$. Clearly, the random vector $\mathbf{z}$ is a GMRF because it follows an improper multivariate Gaussian distribution and satisfies *Markov conditional independence* assumptions. The conditional dependence structure and coefficients of each interior z_{jk} are shown in Figure 1. Rue and Held (2005) introduced (2.10) as a restriction of an infinite intrinsic GMRF to the finite lattice of interest without correcting for the boundaries.

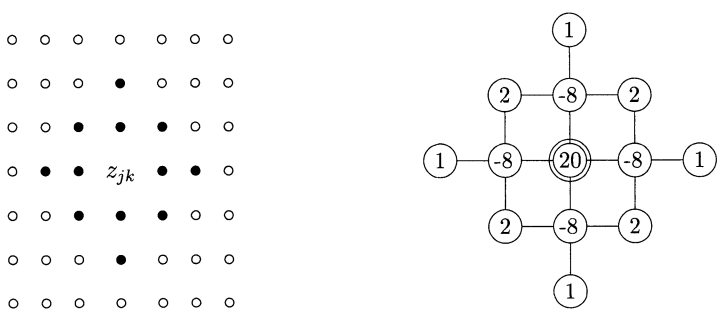

Figure 1. Dependence structure and coefficients of the structure matrix $\mathbf{A}_0$ for the bivariate second-order GMRF in (2.10): the twelve next neighbors of z_{jk} are indicated as black points.

Whereas it may be used with MCMC simulation techniques, (2.8) is not attractive in a Bayesian thin-plate spline setting because it enlarges the null space of the structure matrix $\mathbf{A}_1$ up to $2(n_1 + n_2 - 2)$ dimensions and leads to a noninterpretable basis of this null space (Kneib 2006). To fix this rank-deficiency issue, we add the following boundary terms. At the four corner points we add differential operators of the form $(\pm\frac{\partial}{\partial\tilde{u}} \pm \frac{\partial}{\partial\tilde{v}})$. The sign depends on the boundary, with a positive term at $\tilde{u}_2$ or $\tilde{v}_2$ and a negative term at $\tilde{u}_{n_1}$ or $\tilde{v}_{n_2}$. Similarly, along the edges, we add differential operators of the form $(\frac{\partial^2}{\partial\tilde{u}^2} \pm \frac{\partial}{\partial\tilde{v}})$ or $(\pm\frac{\partial}{\partial u} + \frac{\partial^2}{\partial v^2})$, with the same convention in signs as in the first-order terms. With $\nabla_{(1,0)} f(\tilde{u}_j, \tilde{v}_k) = f(\tilde{u}_j, \tilde{v}_k) - f(\tilde{u}_{j-1}, \tilde{v}_k)$ and $\nabla_{(0,1)} f(\tilde{u}_j, \tilde{v}_k) = f(\tilde{u}_j, \tilde{v}_k) - f(\tilde{u}_j, \tilde{v}_{k-1})$, we introduce eight new operators:

$$\begin{aligned}
\nabla_1 f(\tilde{u}_1, \tilde{v}_1) &\triangleq \big(\nabla_{(1,0)} + \nabla_{(0,1)}\big) f(\tilde{u}_2, \tilde{v}_2),\\
\nabla_2 f\big(\tilde{u}_{n_1}, \tilde{v}_1\big) &\triangleq -\nabla_{(1,0)} f\big(\tilde{u}_{n_1}, \tilde{v}_1\big) + \nabla_{(0,1)} f\big(\tilde{u}_{n_1}, \tilde{v}_2\big),\\
\nabla_3 f\big(\tilde{u}_1, \tilde{v}_{n_2}\big) &\triangleq \nabla_{(1,0)} f\big(\tilde{u}_2, \tilde{v}_{n_2}\big) - \nabla_{(0,1)} f\big(\tilde{u}_1, \tilde{v}_{n_2}\big),\\
\nabla_4 f\big(\tilde{u}_{n_1}, \tilde{v}_{n_2}\big) &\triangleq -\big(\nabla_{(1,0)} + \nabla_{(0,1)}\big) f\big(\tilde{u}_{n_1}, \tilde{v}_{n_2}\big),\\
\nabla_5 f(\tilde{u}_j, \tilde{v}_1) &\triangleq \nabla^2_{(1,0)} f(\tilde{u}_{j+1}, \tilde{v}_1) + \nabla_{(0,1)} f(\tilde{u}_j, \tilde{v}_2),\\
\nabla_6 f\big(\tilde{u}_j, \tilde{v}_{n_2}\big) &\triangleq \nabla^2_{(1,0)} f\big(\tilde{u}_{j+1}, \tilde{v}_{n_2}\big) - \nabla_{(0,1)} f\big(\tilde{u}_j, \tilde{v}_{n_2}\big),\\
\nabla_7 f(\tilde{u}_1, \tilde{v}_k) &\triangleq \nabla_{(1,0)} f(\tilde{u}_2, \tilde{v}_k) + \nabla^2_{(0,1)} f(\tilde{u}_1, \tilde{v}_{k+1}),\\
\nabla_8 f\big(\tilde{u}_{n_1}, \tilde{v}_k\big) &\triangleq -\nabla_{(1,0)} f\big(\tilde{u}_{n_1}, \tilde{v}_k\big) + \nabla^2_{(0,1)} f\big(\tilde{u}_{n_1}, \tilde{v}_{k+1}\big).
\end{aligned} \tag{2.11}$$

Using the operators defined in (2.7) and (2.11), an improved approximation of the roughness penalty in (2.5) can be expressed as

$$\begin{aligned}
\frac{1}{h^4}\Bigg[&\sum_{j=2}^{n_1-1}\sum_{k=2}^{n_2-1}\{\nabla_0^2 f(\tilde{u}_j, \tilde{v}_k)\}^2 + \{\nabla_1 f(\tilde{u}_1, \tilde{v}_1)\}^2 + \big\{\nabla_2 f\big(\tilde{u}_{n_1}, \tilde{v}_1\big)\big\}^2\\
&+ \big\{\nabla_3 f\big(\tilde{u}_1, \tilde{v}_{n_2}\big)\big\}^2 + \big\{\nabla_4 f\big(\tilde{u}_{n_1}, \tilde{v}_{n_2}\big)\big\}^2 + \sum_{j=2}^{n_1-1}\Big(\{\nabla_5 f(\tilde{u}_j, \tilde{v}_1)\}^2 + \big\{\nabla_6 f\big(\tilde{u}_j, \tilde{v}_{n_2}\big)\big\}^2\Big)\\
&+ \sum_{k=2}^{n_2-1}\Big(\{\nabla_7 f(\tilde{u}_1, \tilde{v}_k)\}^2 + \big\{\nabla_8 f\big(\tilde{u}_{n_1}, \tilde{v}_k\big)\big\}^2\Big)\Bigg].
\end{aligned} \tag{2.12}$$

Although the notation for these extra boundary terms is somewhat confusing, the operator in (2.12) has a symmetric representation and can be written as $\mathbf{z}'\tilde{\mathbf{A}}\mathbf{z}$ again with notation $z_{jk} = f(\tilde{u}_j, \tilde{v}_k)$. The structure matrix $\tilde{\mathbf{A}}$ has the coefficients as shown in Figure 2. With the proposed edge corrections, the null space of $\tilde{\mathbf{A}}$ is now one dimensional and spanned by a constant vector. More importantly, a simple factorization $\tilde{\mathbf{A}} = \tilde{\mathbf{B}}'\tilde{\mathbf{B}}$ exists, where $\tilde{\mathbf{B}}$ is an $n \times n$ sparse matrix specified by the difference operators used in (2.12). However, $\tilde{\mathbf{B}}$ has rank $n - 1$ (not full rank). Because full-rank factorization is a key to the success of our

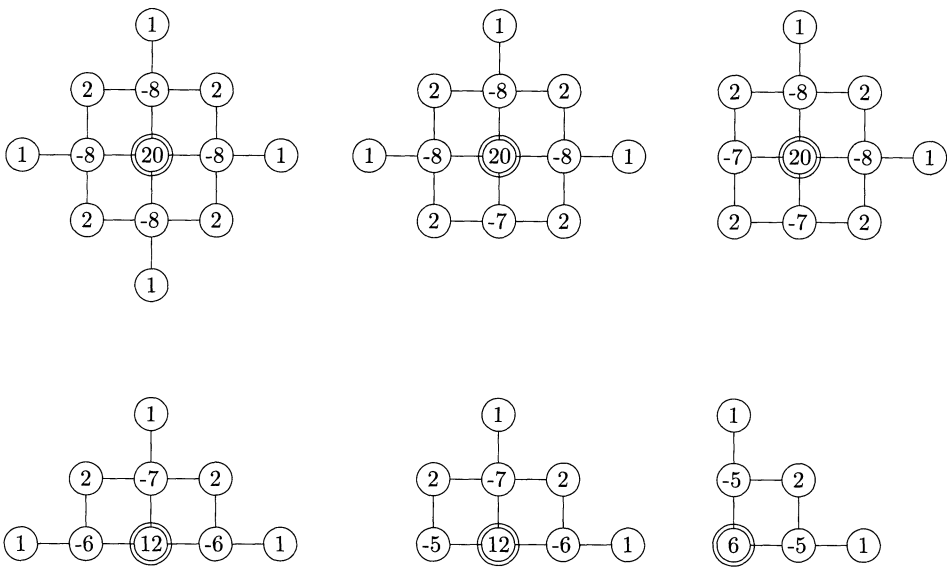

Figure 2. Dependence structure and coefficients of structure matrix $\tilde{\mathbf{A}}$ with boundary corrections.

modeling, we delete the first row of $\tilde{\mathbf{B}}$ to obtain an $(n-1) \times n$ matrix $\mathbf{B}$ whose rank is $n-1$. That $\mathbf{B}$ has full rank is a consequence of the following fact.

Fact 1. *Suppose* $\tilde{\mathbf{B}}$ *is an* $n \times n$ *matrix with rank* $r(\tilde{\mathbf{B}}) = n-1$. *Let* $\mathbf{B}$ *be the* $(n-1) \times n$ *matrix obtained by deleting the first row of* $\tilde{\mathbf{B}}$. *Then* $\mathbf{B}$ *is full rank, that is,* $r(\mathbf{B}) = n-1$.

Proof: Define an $(n-1) \times n$ matrix $\mathbf{E} = (\mathbf{0}, \mathbf{I}_{n-1})$, where $\mathbf{0}$ is a vector of $n-1$ zeroes and $\mathbf{I}_{n-1}$ is an identity matrix of dimension $n-1$. Then, we have $\mathbf{B} = \mathbf{E}\tilde{\mathbf{B}}$. Because $r(\mathbf{B}) = r(\mathbf{B}\mathbf{B}')$, it is sufficient to prove that matrix $\mathbf{A} = \mathbf{B}\mathbf{B}'$ is full rank, that is, $r(\mathbf{A}) = n-1$. Letting $\tilde{\mathbf{A}} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}'$, we have $\mathbf{A} = \mathbf{E}\tilde{\mathbf{A}}\mathbf{E}'$, that is, $\mathbf{A}$ can be obtained by deleting the first row and first column from $\tilde{\mathbf{A}}$. Because $r(\tilde{\mathbf{A}}) = r(\tilde{\mathbf{B}}) = n-1$, $\mathbf{A}$ is full rank by Brezger, Fahrmeir, and Hennerfeind (2007). □

In theory, one can delete any row of $\tilde{\mathbf{B}}$ to make $\mathbf{B}$ full rank. Removing the first row minimizes the subsequent effect on the conditional independence structure of $\mathbf{z}$ because only the upper-left corner point and its five neighbors are affected. The resulting prior lacks some symmetry because we only delete one corner point, but we have not seen any artifacts in our experience with real or simulated datasets.

Letting $\mathbf{A} = \mathbf{B}'\mathbf{B}$, the proposed prior for $\mathbf{z}$ thus has analogous form with (2.10),

$$[\mathbf{z}|\delta] \propto \delta^{(n-1)/2} |\mathbf{A}|_+^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}\mathbf{z}\right), \tag{2.13}$$

where the null space of $\mathbf{A}$ has an interpretable constant vector basis. We call (2.13) a modified second-order intrinsic GMRF in the sense that $\mathbf{z}$ has a second-order spatial dependence structure over the interior, but it is modified on the edge where it is of first order. With the

independent normal error terms $\varepsilon_{jk\ell}$ in (2.2), the posterior distribution of $\mathbf{z}$ can be shown to be $\mathrm{N}_n(\mathbf{S}_{\lambda_h}\mathbf{D}'\mathbf{y}, \tau^{-1}\mathbf{S}_{\lambda_h})$, where the smoothing parameter is $\lambda_h = \delta/\tau$, and the smoother matrix is $\mathbf{S}_{\lambda_h} = (\mathbf{W} + \lambda_h\mathbf{A})^{-1}$ with $\mathbf{W} = \mathbf{D}'\mathbf{D}$. Note that $\mathbf{W}$ is a diagonal matrix whose entries are the numbers of observations at each grid point, and $\mathbf{W} + \lambda_h\mathbf{A}$ must be invertible for the posterior to exist. The posterior mean $\mathbf{S}_{\lambda_h}\mathbf{D}'\mathbf{y}$ is a discretized thin-plate spline estimator. After specifying the priors on precision parameters τ and δ, one can carry out fully Bayesian inference by Gibbs sampling. One should note that there do exist some other bivariate GMRFs that have successful applications in different scenarios (see, e.g., Terzopoulos 1988; Rue and Held 2005; Kneib 2006), but none of them appear to have the full-rank factorization we need for the efficient nonstationary extension in the next section.

2.3 NONSTATIONARY SPATIAL GMRF PRIORS

An alternative representation of (2.13) in the context of state space modeling is

$$\mathbf{Bz} \sim \mathrm{N}_{n-1}(\mathbf{0}, \delta^{-1}\mathbf{I}). \tag{2.14}$$

Here the differences in (2.12) except for ∇_1 are identically independently distributed Gaussian with mean zero and variance δ^{-1}. As suggested by Lang, Fronk, and Fahrmeir (2002), Brezger, Fahrmeir, and Hennerfeind (2007), and Yue, Speckman, and Sun (2008), we replace the constant precision δ with $(n-1)$ locally varying precisions δ_{jk} to achieve an adaptive extension of (2.14). Clearly, the δ_{jk} are expected to be large for the flat areas of the function of interest, whereas they should be small for the highly changing regions.

A further prior must be taken on the δ_{jk} to complete the nonstationary spatial GMRF. Let $\delta_{jk} = \delta e^{\gamma_{jk}}$, where δ is a positive scale parameter and a linear constraint $\sum_{j,k}\gamma_{jk} \equiv 0$ is used for identifiability. We assume the γ_{jk} are smooth as well, and the prior on $\boldsymbol{\gamma} = \mathrm{Vec}([\gamma_{jk}]) = (\gamma_{21}, \gamma_{31}, \ldots, \gamma_{n_1n_2})'$ is a first-order intrinsic GMRF on a regular lattice (Besag and Higdon 1999; Rue and Held 2005) subject to the constraint

$$[\boldsymbol{\gamma}|\eta] \propto \eta^{(n-2)/2}\exp\left\{-\frac{\eta}{2}\left[\sum_{j=3}^{n_1}\sum_{k=2}^{n_2}(\gamma_{jk}-\gamma_{j-1,k})^2 + \sum_{j=2}^{n_1}\sum_{k=3}^{n_2}(\gamma_{jk}-\gamma_{j,k-1})^2 + \sum_{j=3}^{n_1}(\gamma_{j1}-\gamma_{j-1,1})^2 + \sum_{k=3}^{n_2}(\gamma_{1k}-\gamma_{1,k-1})^2\right]\right\} I_{(\mathbf{1}'\boldsymbol{\gamma}=0)}. \tag{2.15}$$

Note that prior (2.15) simply extends the difference approximation for the first derivative by taking into account both horizontal and vertical directions, and the neighborhood structure in this situation is based on the four nearest neighbors. Theoretically, one can take any reasonable latent spatial process prior for $\boldsymbol{\gamma}$, for example, a conditional autoregressive process or a stationary Gaussian process. For consistency, simplicity, and efficient computation, we choose the first-order bivariate GMRF in (2.15).

The priors on $\mathbf{z}$ and $\boldsymbol{\gamma}$ can now be written in matrix notation as

$$[\mathbf{z}|\delta, \boldsymbol{\gamma}] \propto \delta^{(n-1)/2}|\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2}\exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}\right), \tag{2.16}$$

where $\mathbf{A}_{\boldsymbol{\gamma}} = \mathbf{B}'\boldsymbol{\Lambda}_{\boldsymbol{\gamma}}\mathbf{B}$ is an adaptive structure matrix with $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}} = \operatorname{diag}(e^{\gamma_{21}}, \ldots, e^{\gamma_{n_1 n_2}})$, and

$$[\boldsymbol{\gamma}|\eta] \propto \eta^{(n-2)/2} \exp\left(-\frac{\eta}{2}\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}\right) I_{(\mathbf{1}'\boldsymbol{\gamma}=0)}, \tag{2.17}$$

where the null space of $\mathbf{M}$ is spanned by the constant vector $\mathbf{1}$. We refer to (2.16) and (2.17) together as a nonstationary spatial GMRF prior. The proposed nonstationary GMRF prior has appealing properties for Bayesian inference. First, it improves surface smoothing because the variance function is adaptive to changing smoothness of the spatial regions of interest. Second, one can easily implement a Gibbs sampler for Bayesian inference, and the sparseness of the prior yields an efficient MCMC simulation (see Section 3 for details). Third, there is no need to compute $|\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2}$ in (2.16) because $|\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2} = |\mathbf{B}\mathbf{B}'|^{1/2}$ is constant (see Yue, Speckman, and Sun 2008). This identity overcomes the main problem of using a GMRF prior in the second hierarchy to build an adaptive GMRF in the article by Brezger, Fahrmeir, and Hennerfeind (2007). Finally, the modified GMRF prior is second order in the interior, which yields more smoothness and flexibility for function estimation than the weighted first-order GMRF in the article by Brezger, Fahrmeir, and Hennerfeind (2007).

2.4 Hyperpriors and Posterior Propriety

Hyperpriors on the precision components τ, δ, and η are required. Statisticians often take relatively diffuse inverse-gamma priors on the variance components in random effects linear models, or equivalently gamma priors on the precision components. However, it is well known that limiting diffuse inverse-gamma priors lead to improper posterior distributions (e.g., Hobert and Casella 1996). The same problem holds in the case of nonadaptive thin-plate spline types of GMRF priors (Speckman and Sun 2003). Moreover, even in the nonadaptive case, a close reading of the article by Speckman and Sun (2003) suggests that there are no universally good choices of noninformative extended gamma-type priors on τ and δ. On the other hand, there seems to be no way to elicit reasonable prior information on the precision parameter δ for the GMRF $\mathbf{z}$ in the nonadaptive case, and δ effectively controls the amount of smoothing of the data. The problem of prior specification in the adaptive case is even harder.

Because of the infinite-dimensional nature of the smoothing problem, we believe that some prior information must be supplied or inference is impossible. Our philosophy is to introduce this information as deep in the hierarchy as possible. Following Yue, Speckman, and Sun (2008), we believe it makes most sense to first reparameterize the precision components in terms of "smoothing parameters." Specifically, let $\xi_1 = \delta/\tau$ and $\xi_2 = \eta/\delta$. This establishes a one-to-one transformation between (τ, ξ_1, ξ_2) and (τ, δ, η). The parameters ξ_1 and ξ_2 are equivalent to smoothing parameters in the thin-plate spline setup and are also related to Zellner's g-prior (Zellner and Siow 1980; Zellner 1986). The nonstationary spatial GMRF prior then becomes

$$\begin{aligned}
[\mathbf{z}|\tau, \xi_1, \boldsymbol{\gamma}] &\propto (\tau\xi_1)^{(n-1)/2}|\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2} \exp\left(-\frac{\tau\xi_1}{2}\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}\right), \\
[\boldsymbol{\gamma}|\tau, \xi_1, \xi_2] &\propto (\tau\xi_1\xi_2)^{(n-2)/2} \exp\left(-\frac{\tau\xi_1\xi_2}{2}\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}\right) I_{(\mathbf{1}'\boldsymbol{\gamma}=0)}.
\end{aligned} \tag{2.18}$$

Following Jeffreys and Zellner, Yue, Speckman, and Sun (2008) chose the invariance prior on the precision component of the error term, τ, so we also take $[\tau] \propto 1/\tau, \tau > 0$. In the nonadaptive problem, results of Sun and Speckman (2008) imply that a proper prior on ξ_1 must be used to ensure a proper posterior. Thus the problem is in eliciting prior information for proper priors on the smoothing parameters ξ_1 and ξ_2. In the next subsection, we demonstrate how it is possible to elicit prior information on the two smoothing parameters using the notion of "equivalent degrees of freedom."

The following theoretical result helps guide the choice of prior on (ξ_1, ξ_2). Recall that $\mathbf{W} = \mathbf{D}'\mathbf{D}$ and define $\text{SSE} = \mathbf{y}'(\mathbf{I}_N - \mathbf{D}\mathbf{W}^{-}\mathbf{D}')\mathbf{y}$, where $\mathbf{W}^{-}$ is a generalized inverse of $\mathbf{W}$. We then consider the following cases:

Case 1: $\text{SSE} > 0$.

Case 2: $\text{SSE} = 0$ and $\mathbf{W} = \mathbf{I}_N$.

Case 3: $\text{SSE} = 0$ and $\mathbf{W}$ is positive semidefinite.

In Case 1, we have repeated measurements. In Case 2, there is a single observation at each location. In Case 3, there is at most one observation for each location, and some locations have missing data. With $[\tau] \propto 1/\tau$, the following theorem provides sufficient conditions on the possibly joint prior $[\xi_1, \xi_2]$ for a proper posterior distribution in each of the three cases.

Theorem 1. *Consider the nonparametric model* (2.2) *with prior distribution given by* (2.18) *and* $[\tau] \propto 1/\tau$, *and assume the following conditions*:

Case 1: (a) $[\xi_1, \xi_2]$ *is proper.*

Case 2: (a) $[\xi_1, \xi_2]$ *is proper*; (b) $E\xi_2^{-(N-1)/2} < \infty$.

Case 3: (a) $[\xi_1, \xi_2]$ *is proper*; (b) $E(\xi_1\xi_2)^{-(N-1)/2} < \infty$.

Then the joint posterior of $(\mathbf{z}, \tau, \xi_1, \xi_2, \boldsymbol{\gamma})$ *is proper.*

The proof is in the supplemental materials. Following Yue, Speckman, and Sun (2008), we suggest independent Pareto priors for ξ_1 and ξ_2 if the prior is only required to be proper, and inverse-gamma priors if priors must have finite negative moments. For example, in Case 2 we take $[\xi_1|c] = c/(c+\xi_1)^2, \xi_1 > 0, c > 0$, and $[\xi_2|a, b] \propto \xi_2^{-(a+1)}e^{-b/\xi_2}$, $\xi_2 > 0, a > 0, b > 0$.

2.5 HYPERPARAMETERS AND EQUIVALENT DEGREES OF FREEDOM

The results of Theorem 1 suggest that informative priors must be used for ξ_1 and ξ_2. However, it appears difficult to elicit hyperparameters for $[\xi_1|c]$ and $[\xi_2|a, b]$ directly because ξ_1 and ξ_2 are also not readily interpretable. Instead, following an idea of Hastie, Tibshirani, and Friedman (2001), and adapted by White (2006), we choose the hyperparameters to satisfy certain "equivalent degrees of freedom" conditions on smoother matrices or approximate smoother matrices. The development here is taken from the work of Yue, Speckman, and Sun (2008).

Again we take Case 2 as an example. Letting $\boldsymbol{\gamma} = \mathbf{0}$, the smoother matrix for $\mathbf{z}$ is $\mathbf{S}_{\xi_1} = (\mathbf{W} + \xi_1\mathbf{A})^{-1}$. Because the trace of the smoother matrix is often taken as the degrees of freedom for nonparametric regression (e.g., Hastie, Tibshirani, and Friedman

2001), we choose the value c, which is the median of the Pareto prior on ξ_1, that sets trace($\mathbf{S}_c$) equal to desired prior degrees of freedom. Because trace($\mathbf{S}_{\xi_1}$) is a monotone function of ξ_1, this procedure sets the median of the prior on trace($\mathbf{S}_{\xi_1}$) to the desired amount. Similarly, we compute an equivalent smoother matrix for $\boldsymbol{\gamma}$ using a Laplace approximation to expand its full conditional (see the appendix in Yue, Speckman, and Sun 2008). Letting $\hat{\mathbf{z}}$ be any nonadaptive estimate and $\tilde{\mathbf{z}} = \mathbf{B}\hat{\mathbf{z}}$, the approximate smoother matrix is $(\mathbf{H} + 2\xi_2\mathbf{M})^{-1}\mathbf{H}$, where $\mathbf{H} = \text{diag}(\tilde{z}_1^2, \ldots, \tilde{z}_{n-1}^2)$. The equivalent degrees of freedom is $\text{trace}(\mathbf{H} + 2\xi_2\mathbf{M})^{-1}\mathbf{H} = \text{trace}(\mathbf{I}_{n-1} + 2\xi_2\mathbf{H}^{-1}\mathbf{M})^{-1}$. Fixing shape $a = 0.5$, we typically choose a scale parameter b giving desired degrees of freedom.

The notion of "degrees of freedom" relates directly to the complexity of the model. For a parametric model, the degrees of freedom is exactly the number of parameters in the model. "Equivalent degrees of freedom" has exactly the same interpretation. Thus a researcher can use his or her knowledge of the problem to select appropriate prior distributions for ξ_1 and ξ_2. In Section 4.2, we provide more guidelines on the choice of the prior degrees of freedom. Fortunately, we always observe Bayesian learning, and the choice of prior is not critical to the predicted surface.

3. IMPLEMENTATION USING GIBBS SAMPLER

3.1 FULL CONDITIONAL DISTRIBUTIONS

Recall that we take a nonstationary GMRF on $\mathbf{z}$ and $\boldsymbol{\gamma}$, an invariance prior on τ, a Pareto prior on ξ_1, and an inverse gamma prior on ξ_2. Note that the Pareto prior on ξ_1 can be seen to be a scale mixture of exponentials by introducing a latent variable θ, that is, $[\xi_1|c] = \int [\xi_1|\theta][\theta|c]\,d\theta$, where $[\xi_1|\theta] = \theta\exp(-\theta\xi_1)$ and $[\theta|c] = c\exp(-c\theta)$. The full conditionals of the posterior distribution for Case 2 are detailed below.

1. $(\mathbf{z}|\mathbf{y}, \xi_1, \tau, \boldsymbol{\gamma}) \sim \text{N}_n(\mathbf{S}_{\xi_1}\mathbf{D}'\mathbf{y}, \tau^{-1}\mathbf{S}_{\xi_1})$, where $\mathbf{S}_{\xi_1} = (\mathbf{W} + \xi_1\mathbf{A}_{\boldsymbol{\gamma}})^{-1}$.
2. $[\boldsymbol{\gamma}|\mathbf{z}, \xi_1, \xi_2] \propto \exp(-\tau\xi_1\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}/2 - \tau\xi_1\xi_2\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}/2)I_{(\mathbf{1}'\boldsymbol{\gamma}=0)}$.
3. $(\tau|\mathbf{y}, \mathbf{z}, \boldsymbol{\gamma}, \xi_1, \xi_2) \sim \text{Gamma}(n + (N-3)/2, \|\mathbf{y} - \mathbf{D}\mathbf{z}\|^2/2 + \xi_1\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}/2 + \xi_1\xi_2\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}/2)$.
4. $(\xi_1|\mathbf{z}, \tau, \xi_1) \sim \text{Gamma}(n - 1/2, \tau\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}/2 + \tau\xi_2\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}/2 + \theta)$.
5. $(\xi_2|\boldsymbol{\gamma}, \xi_1) \propto \xi_2^{(n-2)/2-a-1}\exp(-\tau\xi_1\xi_2\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}/2 - b/\xi_2)$.
6. $(\theta|\xi_1) \sim \text{Gamma}(2, \xi_1 + c)$.

To sample $\mathbf{z}$ as a block, the inverse of matrix $\mathbf{Q} = \mathbf{W} + \xi_1\mathbf{A}_{\boldsymbol{\gamma}}$ must be computed. Note that $\mathbf{Q}$ is an $n \times n$ matrix. Therefore, even for moderate resolution, say $n_1 = n_2 = 30$, solving a 900×900 matrix is computationally intensive. Fortunately, $\mathbf{Q}$ is banded with bandwidth at most $2n_1$ due to the construction of $\mathbf{A}_{\boldsymbol{\gamma}}$. Thus, $\mathbf{Q}$ can be factorized by a band Cholesky decomposition, and the inverse can be obtained efficiently by solving a lower triangle linear system. (The band matrix $\mathbf{Q}$ is also sparse, but further attempts to reduce the bandwidth using the Gibbs–Poole–Stockmeyer reorder algorithm gave little improvement.) The full

conditional of ξ_2 is log concave, and hence the adaptive rejection Metropolis sampling (ARMS) (Gilks and Wild 1992; Gilks, Best, and Tan 1995) is applicable for sampling ξ_2. To speed the convergence of Markov chains, we propose a block-move sampling method for $\boldsymbol{\gamma}$ based on a Metropolis–Hastings algorithm in the next section.

3.2 BLOCK SAMPLING ALGORITHM FOR UPDATING $\boldsymbol{\gamma}$

We sample $\boldsymbol{\gamma}$ by first dividing the vector into several small blocks and then updating each block by a Metropolis–Hastings algorithm with the proposal being the prior taken on that block conditioned on the boundary values. The basic idea was introduced by Carter and Kohn (1996) and Knorr-Held and Richardson (2003), and we now apply it to spatial modeling.

The terms that involve $\boldsymbol{\gamma}$ in the posterior distribution are

$$\exp\left(-\frac{\tau\xi_1}{2}\mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z} - \frac{\tau\xi_1\xi_2}{2}\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}\right) I_{(\mathbf{1}'\boldsymbol{\gamma}=0)}. \tag{3.1}$$

Let $\boldsymbol{\Gamma} = [\gamma_{jk}]$ be the lattice of γ_{jk} as in Figure 3(a). Note that $\boldsymbol{\Gamma}$ is obtained by removing γ_{11} from the $n_1 \times n_2$ matrix of $\gamma_{\ell m}$. We then select the blocks $\boldsymbol{\Gamma}[r_1 : s_1, r_2 : s_2]$ given by the rows numbered r_1 to s_1 and the columns numbered r_2 to s_2, where $1 \le r_1 \le s_1 \le n_1$ and $1 \le r_2 \le s_2 \le n_2$. Several possible blocks are presented in Figure 3(a).

Letting $\boldsymbol{\gamma}_{rs} = \mathrm{Vec}(\boldsymbol{\Gamma}[r_1 : s_1, r_2 : s_2])$, the full conditionals for $\boldsymbol{\gamma}_{rs}$ satisfy

$$[\boldsymbol{\gamma}_{rs}|\boldsymbol{\gamma}_{-rs}, \mathbf{z}, \tau, \xi_1, \xi_2] \propto p(\mathbf{z}|\boldsymbol{\gamma}_{rs}, \tau, \xi_1)\pi^*(\boldsymbol{\gamma}_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2, \mathbf{1}'\boldsymbol{\gamma} = 0).$$

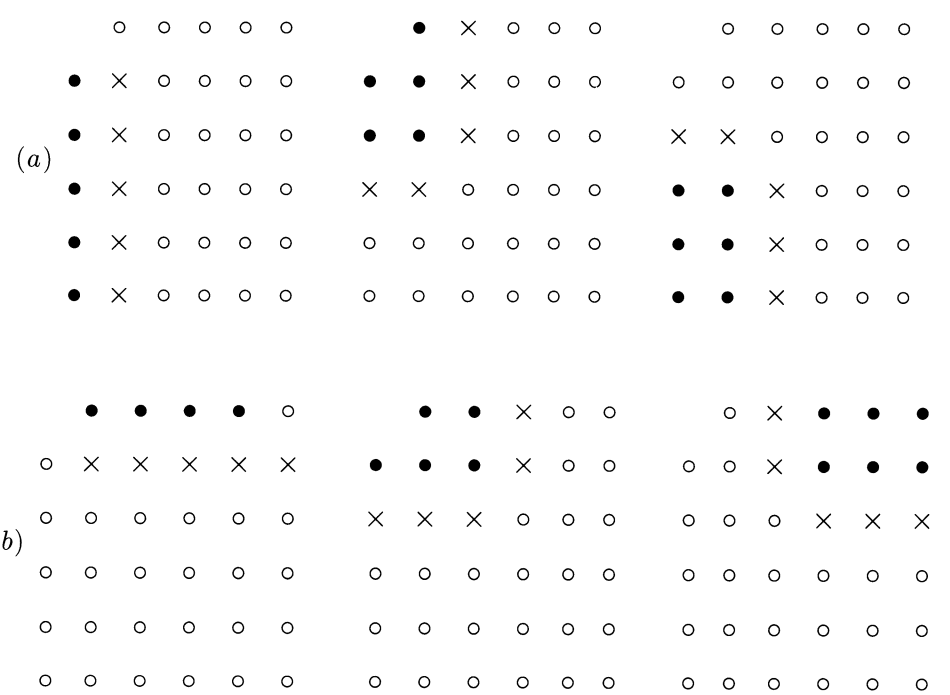

Figure 3. Examples of block $\boldsymbol{\gamma}_{rs}$ in step 1 (2) for $n_1 = n_2 = 6$. Black dots indicate the elements of the block, and the crosses indicate the relevant neighbors.

The first factor $p(\mathbf{z}|\boldsymbol{\gamma}_{rs}, \tau, \xi_1)$ is the product of all terms in the prior density of $\mathbf{z}$ that depend on $\boldsymbol{\gamma}_{rs}$. The second factor is the full conditional of $\boldsymbol{\gamma}_{rs}$ in the prior for $\boldsymbol{\gamma}$, a multivariate Gaussian distribution subject to $\mathbf{1}'\boldsymbol{\gamma} = 0$. Metropolis–Hastings block-move updates for $\boldsymbol{\gamma}_{rs}$ are obtained by drawing a candidate $\boldsymbol{\gamma}^*_{rs}$ from π^* and accepting with probability

$$\min\left(1, \frac{p(\mathbf{z}|\gamma^*_{rs}, \tau, \xi_1)\pi^*(\boldsymbol{\gamma}^*_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2, \mathbf{1}'\boldsymbol{\gamma} = 0)}{p(\mathbf{z}|\gamma_{rs}, \tau, \xi_1)\pi^*(\boldsymbol{\gamma}_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2, \mathbf{1}'\boldsymbol{\gamma} = 0)} \frac{\pi^*(\boldsymbol{\gamma}_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2, \mathbf{1}'\boldsymbol{\gamma} = 0)}{\pi^*(\boldsymbol{\gamma}^*_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2, \mathbf{1}'\boldsymbol{\gamma} = 0)}\right),$$

which is as simple as

$$\min\left(1, \frac{p(\mathbf{z}|\gamma^*_{rs}, \tau, \xi_1)}{p(\mathbf{z}|\gamma_{rs}, \tau, \xi_1)}\right).$$

It is not straightforward to sample $\boldsymbol{\gamma}^*_{rs}$ from the conditional distribution π^*. Our strategy is to first sample $\boldsymbol{\gamma}^{**}_{rs}$ from $\pi^{**}(\boldsymbol{\gamma}_{rs}|\boldsymbol{\gamma}_{-rs}, \tau, \xi_1, \xi_2)$, an unconstrained version of π^*, and then correct $\boldsymbol{\gamma}^{**}_{rs}$ to obtain $\boldsymbol{\gamma}^*_{rs}$ (see Rue and Held 2005, sec. 2.3.3).

Due to the Markov property, the conditional distribution of $\boldsymbol{\gamma}_{rs}$ only depends on the boundary neighbors as shown in Figure 3. We thus define $\tilde{\boldsymbol{\gamma}}_{rs} = (\boldsymbol{\gamma}'_{rs}, \tilde{\boldsymbol{\gamma}}'_{-rs})'$, a vector of the block $\boldsymbol{\gamma}_{rs}$ and its relevant neighbors $\tilde{\boldsymbol{\gamma}}_{-rs}$. It is obvious that $\tilde{\boldsymbol{\gamma}}_{rs}$ is normally distributed with

$$[\tilde{\boldsymbol{\gamma}}_{rs}|\tau, \xi_1, \xi_2] \propto \exp\left[-\frac{\tau\xi_1\xi_2}{2}(\boldsymbol{\gamma}'_{rs} \quad \tilde{\boldsymbol{\gamma}}'_{-rs})'\begin{pmatrix}\mathbf{M}_{rs} & \mathbf{M}_{rt}\\ \mathbf{M}'_{rt} & \mathbf{M}_{-rs}\end{pmatrix}\begin{pmatrix}\boldsymbol{\gamma}_{rs}\\ \tilde{\boldsymbol{\gamma}}_{-rs}\end{pmatrix}\right],$$

where $\mathbf{M}_{rs}$, etc., all have elements from $\mathbf{M}$, and $\mathbf{M}_{rs}$ is always positive definite (see Dai 2008). Given the distribution of $\tilde{\boldsymbol{\gamma}}_{rs}$, it is fairly easy to obtain π^{**}, which is Gaussian with mean $\mu^{**}_{rs} = -\mathbf{M}^{-1}_{rs}\mathbf{M}_{rt}\tilde{\boldsymbol{\gamma}}_{-rs}$ and variance $\Sigma^{**}_{rs} = (\tau\xi_1\xi_2)^{-1}\mathbf{M}^{-1}_{rs}$. Because the Cholesky factor of the banded matrix $\mathbf{M}_{rs}$ can be calculated, the $\boldsymbol{\gamma}^{**}_{rs}$ can be simulated from π^{**} efficiently using back-substitution as described for the z-field. We then compute $\boldsymbol{\gamma}^*_{rs} = \boldsymbol{\gamma}^{**}_{rs} - \mathbf{w}_{rs}(\mathbf{1}'\boldsymbol{\gamma}^{**}_{rs} - \mathbf{1}'\boldsymbol{\gamma}_{rs})$ with $\mathbf{w}_{rs} = \mathbf{M}^{-1}_{rs}\mathbf{1}/\mathbf{1}'\mathbf{M}^{-1}_{rs}\mathbf{1}$; $\boldsymbol{\gamma}^*_{rs}$ has the correct constrained distribution π^*. Notice that $\mathbf{w}_{rs}$ can be computed and stored in advance, leaving only simple linear computation in each MCMC iteration.

Unfortunately, the straightforward Markov chain for sampling blocks $\boldsymbol{\gamma}_{rs}$ is reducible due to the restriction $\mathbf{1}'\boldsymbol{\gamma} = 0$. To see this, let the update of the first block at the mth MCMC iteration be $\boldsymbol{\gamma}^{(m)}_{1s} = (\gamma^m_1, \ldots, \gamma^m_s)'$. We have $\sum_{i=1}^{s}\boldsymbol{\gamma}^{(m)}_i = -\sum_{i=s+1}^{n}\gamma^{(m)}_i$ because of the constraint. In the next iteration, the proposal $\boldsymbol{\gamma}^{*(m+1)}_{1s}$ must satisfy the same constraint, $\sum_{i=1}^{s}\boldsymbol{\gamma}^{*(m+1)}_i = -\sum_{i=s+1}^{n}\gamma^{(m)}_i = \sum_{i=1}^{s}\boldsymbol{\gamma}^{(m)}_i$, that is, the sum of the γ_i for each block update never changes throughout MCMC iterations. We therefore suggest a two-step sampling strategy. In the mth iteration we first update all columnwise blocks $\boldsymbol{\gamma}^{(m)}_{rs}$ as described above, and in the next iteration we sample blocks of exactly the same size by row (see Figure 3(b)). This sampling scheme is convenient for programming, especially when the lattice used is square, because the blocks in the two steps have symmetric full conditional distributions.

4. SIMULATION STUDIES

4.1 COMPARISONS WITH OTHER METHODS

We performed a simulation study using two mean surface functions:

1. $f_1(u, v) = 1.9[1.35 + \exp(u)\sin\{13(u - 0.6)^2\}\exp(-v)\sin(7v)]$, $u, v \in (0, 1)$,
2. $f_2(u, v) = 2\exp\{-\frac{1}{0.4}[(u - 2)^2 + (v - 2)^2]\} + \exp\{-\frac{1}{3}(u^2 + v^2)\}$, $u, v \in (-5, 5)$,

with u and v equally spaced in the intervals. Function f_1 is smoothly varying (Holmes and Mallick 2001; Paciorek and Schervish 2004) whereas f_2 with two uneven modes has much more spatial variability. We use a sample size $n = 900$ with $n_1 = n_2 = 30$, $\tau^{-1} = 0.1^2$, and 250 simulations for each function. According to the strategy in Section 2.4, we chose $a = 0.5$ and $b = 0.02$ for f_1 and let $a = 0.5$ and $b = 0.001$ for f_2. These hyperparameters corresponded to approximately 50 and 200 median prior degrees of freedom for ξ_2. For ξ_1 in both cases, we let $c = 8$, corresponding to about 50 median prior degrees of freedom. The block size for sampling $\boldsymbol{\gamma}$ was 10, which resulted in good mixing of Markov chains with acceptances rates between 0.2 and 0.4.

We compared the performance of Bayesian adaptive thin-plate splines (BATS) with that of the fast adaptive P-splines (FAPS) algorithm of Krivobokova, Crainiceanu, and Kauermann (2008), which performed at least as well as or better than other P-splines-based methods including those of Lang and Brezger (2004) and Crainiceanu et al. (2007). To implement FAPS, we used the R package "AdaptFit" and chose 12×12 and 6×6 equally spaced knots for modeling the mean function and adaptive variances, respectively. The performance of all estimators is measured by log(MSE).

The ranges and the three quartiles of log(MSE) from both methods are presented in Table 1. It appears that FAPS slightly outperforms BATS in estimating the slowly varying function f_1, whereas BATS yields a much smaller log(MSE) than FAPS in the bimodal example. We also present typical fits of the bimodal function from the two methods in Figure 4. It appears that FAPS has trouble capturing the sharp peak whereas BATS slightly undersmooths the small bump. We believe this is because FAPS uses P-splines with a limited number of basis elements for local variances, resulting in more smoothness but less adaptivity than the first-order GMRF used in BATS. We also experimented with FAPS using higher dimensional bases. The fit was barely improved, and the algorithm appeared to suffer numerical instability as the dimension grew. Therefore, FAPS seems to be appropriate for slowly varying functions only. To gain more flexibility in BATS, the local γ_k could be assumed to follow a spatial process with a richer spatial dependence structure than

Table 1. Distributions of log (MSE) from simulations.

	f_1: stationary function					f_2: bimodal function				
	Min.	Q_1	Median	Q_3	Max.	Min.	Q_1	Median	Q_3	Max.
FAPS	−6.23	−6.11	−6.07	−6.02	−5.84	−6.41	−6.34	−6.30	−6.26	−6.12
BATS	−6.13	−5.98	−5.91	−5.85	−5.66	−7.40	−7.16	−7.04	−6.95	−6.64

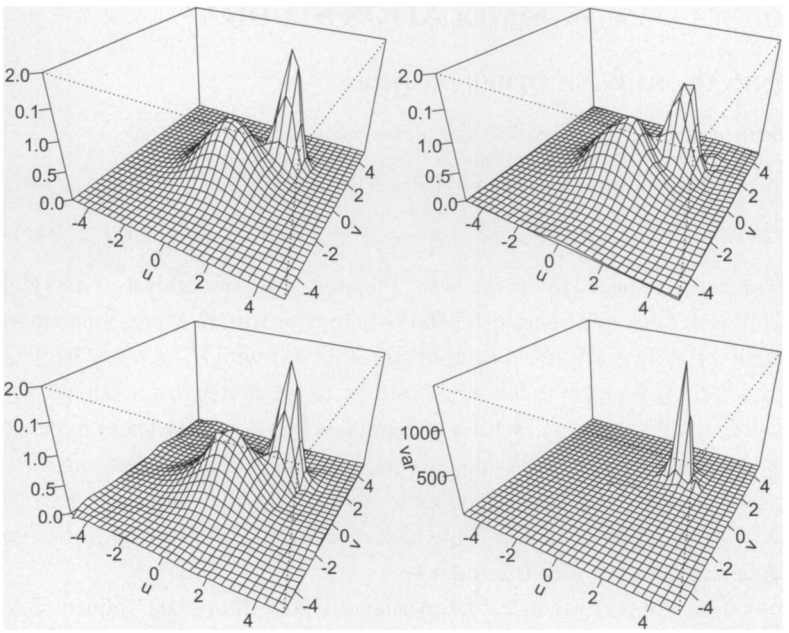

Figure 4. Bimodal simulated example: true surface (top left); fit of FAPS (top right); fit of BATS (bottom left); adaptive variance estimates $e^{-\gamma_{jk}}$ from the BATS fit (bottom right).

the first-order GMRF (2.17), for example, a conditional autoregressive process or Matèrn family.

4.2 Sensitivity Analysis and Choice of Hyperpriors

One needs to specify the hyperparameters for the priors on ξ_1 and ξ_2. As discussed in Section 2.5, we use the notion of equivalent degrees of freedom to choose these hyperparameters, because the smoothing parameters ξ_1 and ξ_2 do not seem to be readily interpretable directly. Median prior degrees of freedom can be chosen subjectively based on knowledge of the problem. Smooth surfaces require a relatively simple fit with fewer degrees of freedom. Complicated surfaces will require more degrees of freedom. In general, fitting a surface requires a large number of degrees of freedom.

Because the model is hierarchical, we use the data to choose appropriate degrees of freedom, guided by the prior distributions. In our experience, the adaptive fit is quite robust to the choice of prior for ξ_1 but somewhat more sensitive to the choice for ξ_2. We experimented with median df $\sim$ 5 and 100 for ξ_1 ($c = 1875$ and 2) when using df $\sim$ 50 for ξ_2 ($b = 0.008$) in the bimodal example and plotted the posterior densities of ξ_1 under both situations in Figure 5 (left). As can be seen, the two densities are fairly similar with great overlapping area. For ξ_2, we tried df $\sim$ 50, 100, 300, and 500 (b =0.008, 0.0029, 0.0004, and 0.0001) with df $\sim$ 50 ($c = 8$) for ξ_1. As shown in the middle panel of Figure 5, the posteriors of ξ_2 when df = 100 and 300 are more separate with distinct posterior means. It is interesting that the overall fit, however, is relatively unaffected by the choice of the df because log(MSE) = -6.93, -7.02, -7.04, and -7.04 when df = 50, 100, 300, and 500,

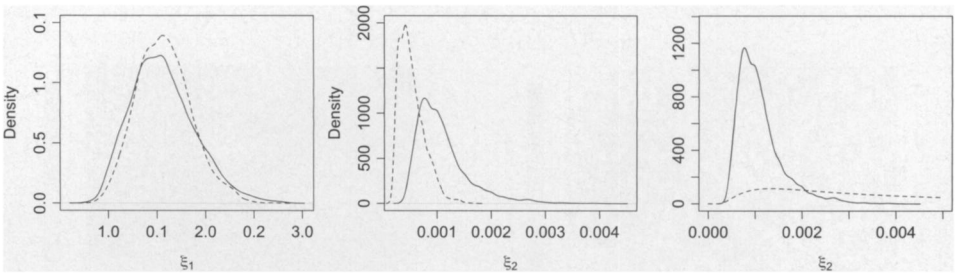

Figure 5. Posterior densities of ξ_1 (left) and ξ_2 (middle) at low prior equivalent df (solid line) and high prior equivalent df (dashed line) for each. Right panel: comparison between posterior density (solid line) and prior density (dashed line) of ξ_2 with the low df prior.

respectively. Moreover, the right panel of Figure 5 shows strong Bayesian learning for the posterior distribution of ξ_2 with the low df prior. The data definitely are the main determinant for the posterior distribution of ξ_2.

In conclusion, the performance of BATS is quite robust to the choice of hyperpriors if $[\xi_2|a, b]$ is chosen to have reasonable df for the approximate smoother on $\boldsymbol{\gamma}$. The value of df can reflect the sample size n and the assumed complexity of the underlying process. We have had success choosing c so that $\mathrm{df}_1 \sim 50$ for ξ_1 and choosing $a = 0.5$ and b such that $\mathrm{df}_2 \sim \min(100, n/3)$ for ξ_2. We may make adjustments to df_2 when a priori information suggests especially smooth or rough functions. In the simulated examples ($n = 900$), we chose $b = 0.02$ for the slowly varying f_1 giving $\mathrm{df}_2 \sim 50$ and $b = 0.001$ for the highly varying f_2 yielding $\mathrm{df}_2 \sim 200$.

4.3 COMPUTATIONAL PERFORMANCE

The BATS method provides reliable and efficient MCMC simulation. Because block sampling is used to update $\boldsymbol{\gamma}$, the Markov chains have a quick convergence: 15,000 MCMC iterations with a burn-in of 5000 proved to be sufficient for all the simulated examples. Due to the sparsity of the GMRF priors, the MCMC computation is fairly fast considering the large number of parameters to be estimated in BATS. With $n = 20^2$, 30^2, and 40^2, a FORTRAN program of BATS took 38, 123, and 390 sec, respectively, for 15,000 iterations on a 2.13 GHz Intel processor. FAPS was faster: it took 26, 42, and 69 sec, respectively. However, FAPS often failed to converge with larger sample sizes, for example, $n = 50^2$ or 60^2. We suspect it is due to the small error variance (0.1^2) used in the simulated examples. In contrast, the BATS MCMC algorithm always yielded stable inference even for large n.

5. APPLICATION TO U.S. PRECIPITATION DATA

The data are annual precipitation measurements from 2082 stations irregularly spaced between latitude 27.1 to 49.0 and longitude -100.5 to -80.2. The time period is from 1960 to 1999. A detailed description of the dataset can be found in the work of Groisman (2000). We chose data for the year 1960 and took the square root of the observation for each station to stabilize the variance. The transformed data are shown in Figure 6, reflecting

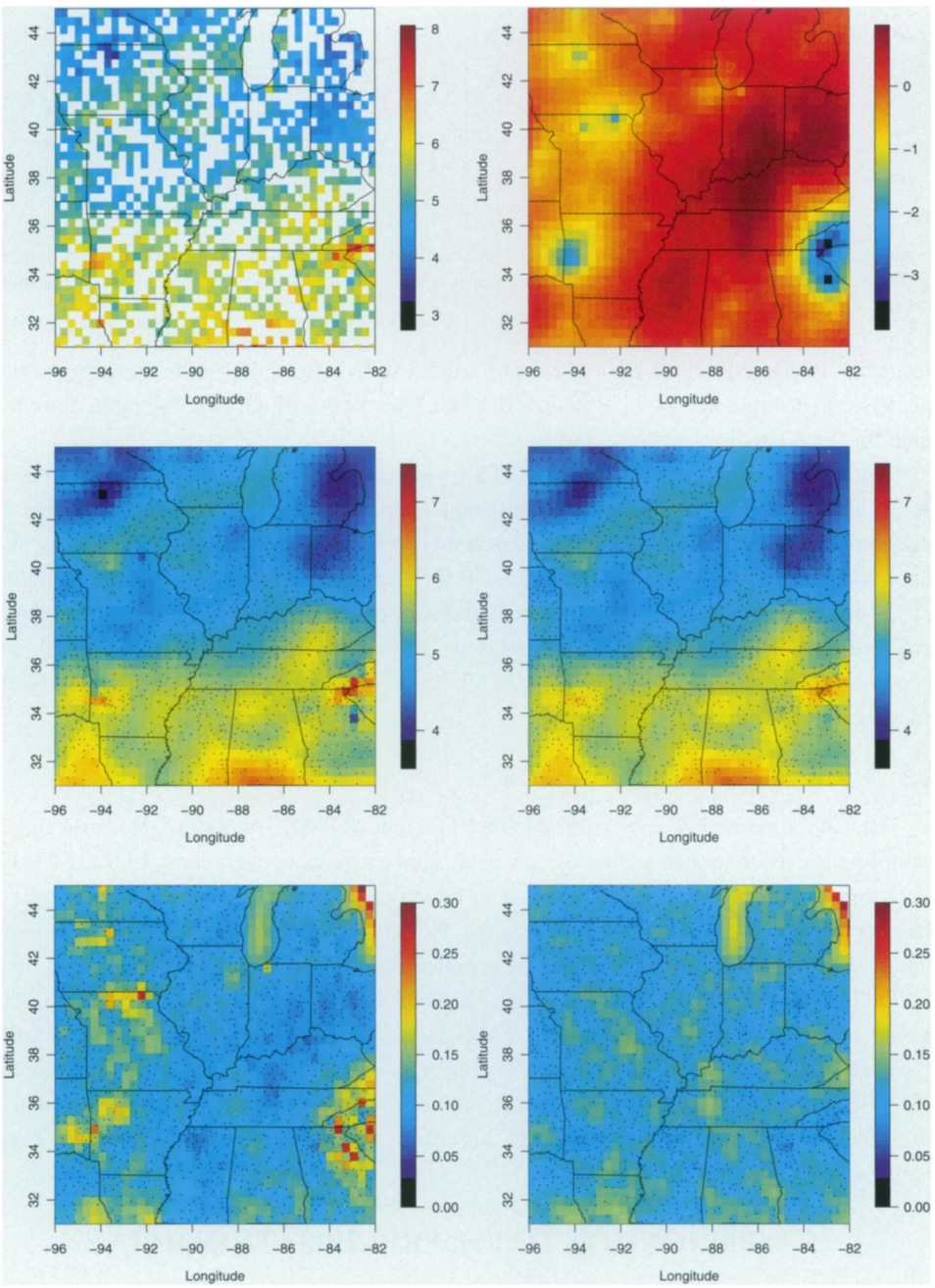

Figure 6. Plot of transformed rainfall data (top left); fit of BATS (middle left); posterior standard deviation of BATS fit (bottom left); estimated γ (top right); fit of TPS (middle right); posterior standard deviation of TPS fit (bottom right).

the general pattern of the precipitation in 1960: in general there was higher rainfall in the south, especially along the boundaries shared by North Carolina, South Carolina, and Georgia, than in the north. Zhu and Wu (2010) analyzed the same dataset but modeled the differences between the observations and the mean over the entire 40-year period.

We binned the data to a 60×60 regular lattice. Fifty-seven percent of the grid points had no data values, and twelve percent of the grid points had multiple observations. As we shall see, the BATS model handles the missing observations well. We chose $c = 1000$ for $[\xi_1|c]$ and $a = 0.5, b = 0.01$ for $[\xi_2|a, b]$, resulting in equivalent degrees of freedom around 100 and 500, respectively; we also picked block size equal to 20 to sample $\boldsymbol{\gamma}$, with resulting acceptance rates between 0.2 and 0.3. We ran the model for 200,000 iterations with 100,000 for burn-in (about 12 hours in FORTRAN for 2028 observations and 3600 prediction locations) and saved every tenth iteration. To compare, we also fit a (nonadaptive) thin-plate spline (TPS) model under prior (2.13). The noninformative gamma priors (Speckman and Sun 2003) are taken on the precisions of TPS, and we ran 20,000 iterations with 10,000 for burn-in (about half an hour computation time).

To illustrate the effect of the spatial inhomogeneity in the rainfall data, we compared the performances of the BATS and TPS models. Figure 6 shows the posterior means and pointwise posterior standard deviations of the spatial surfaces from the two models. The results match our expectation with the features of this area. The surface from BATS is more peaky in the two mountainous regions (Boston Mountains in Arkansas and Great Smoky Mountains in North Carolina) and slightly lower in the northeast (Indiana and Ohio) than the surface from TPS, whereas both are quite smooth in the rest of the study region. The differences between posterior standard deviations are more dramatic: the BATS model gives higher posterior standard deviations in the southeast (west of North Carolina and South Carolina, and north of Georgia) and midwest (west of Arkansas and Missouri), and lower posterior standard deviation in the east (Ohio, Indiana, and Kentucky) and south (Mississippi). In conclusion, the BATS model captures the higher variations of the processes in the mountainous regions compared to the other regions, whereas the TPS model is not able to do so. An image plot of the estimated $\boldsymbol{\gamma}$ is also shown in Figure 6, allowing us to see how the BATS model works: the low $\boldsymbol{\gamma}$ in the mountainous regions brings appropriately large variances to the estimated surface over those regions.

6. SUMMARY AND FUTURE WORK

In this article, we introduce a novel bivariate adaptive GMRF by using a spatially varying variance component with its own GMRF prior. The result is a Bayesian hierarchical model corresponding to a discretized adaptive thin-plate spline suitable for nonstationary spatial data. Bayesian computation is based on an efficient Gibbs sampler, in which the band Cholesky decomposition and a block-move Metropolis–Hastings algorithm are used. The strong performance of our nonstationary spatial model has been demonstrated by two simulated examples and one application.

Given covariates and unknown parameters, the spatial model in (2.1) can be easily extended to the following additive model:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + g_1(t_{i1}) + \cdots + g_r(t_{ir}) + f(u_i, v_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \tau^{-1}), \tag{6.1}$$

where $\mathbf{x}_i'\boldsymbol{\beta}$ is the linear term, $g_1, \ldots, g_r$ are univariate smooth functions, and the bivariate function $f(u_i, v_i)$ represents the spatial effect in the model. To implement Bayesian inference for model (6.1), one can take a diffuse normal or constant prior on $\boldsymbol{\beta}$, univariate adaptive GMRFs (Lang, Fronk, and Fahrmeir 2002; Yue, Speckman, and Sun 2008) or other Gaussian processes for $g_1, \ldots, g_r$, and the nonstationary GMRF proposed here for the spatial function f. This additive model is quite flexible and its estimation can be easily carried out by an efficient Bayesian backfitting algorithm. In addition, the FORTRAN program for (6.1) can be obtained by modifying the one for BATS with small effort.

SUPPLEMENTAL MATERIALS

Appendix: The detailed proof of Theorem 1 for posterior propriety. (appendix.pdf, .pdf file)

Program package: Package containing FORTRAN and R code to perform the BATS model described in the article. (bats-software.tar.gz, zipped tar file)

Precipitation dataset: Dataset used in Section 5. (rain.dat, text file)

ACKNOWLEDGMENT

This research has been supported in part by NSF grant SES-0351523, NIMH grant R01-MH071418, and PSC-CUNY Research Award 60147-39 40.

[Received September 2008. Revised September 2009.]

REFERENCES

Baladandayuthapani, V., Mallick, B. K., and Carroll, R. J. (2005), "Spatially Adaptive Bayesian Penalized Regression Splines (P-Splines)," *Journal of Computational and Graphical Statistics*, 14, 378–394. [97]

Besag, J., and Higdon, D. (1999), "Bayesian Analysis of Agricultural Field Experiments" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 61, 691–746. [103]

Bookstein, F. L. (1989), "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 11, 567–585. [99]

Brezger, A., Fahrmeir, L., and Hennerfeind, A. (2007), "Adaptive Gaussian Markov Random Fields With Applications in Human Brain Mapping," *Journal of the Royal Statistical Society, Ser. C*, 56, 327–345. [97,102-104]

Carter, C. K., and Kohn, R. (1996), "Markov Chain Monte Carlo in Conditionally Gaussian State Space Models," *Biometrika*, 83, 589–601. [107]

Crainiceanu, C., Ruppert, D., Carroll, R., Adarsh, J., and Goodner, B. (2007), "Spatially Adaptive Penalized Splines With Heteroscedastic Errors," *Journal of Computational and Graphical Statistics*, 16, 265–288. [97, 109]

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley-Interscience. [96]

Dai, L. (2008), "Topics in Objective Bayesian Methodology and Spatial-Temporal Models," Ph.D. thesis, University of Missouri-Columbia, Columbia, Missouri. [108]

Damian, D., Sampson, P., and Guttorp, P. (2001), "Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structure," *Environmetrics*, 12, 161–178. [98]

Di Matteo, I., Genovese, C. R., and Kass, R. E. (2001), "Bayesian Curve-Fitting With Free-Knot Splines," *Biometrika*, 88, 1055–1071. [98]

Fuentes, M., and Smith, R. L. (2001), "A New Class of Nonstationary Spatial Models," technical report, North Carolina State University, Dept. of Statistics, available at *www.stat.unc.edu/postscript/rs/nonstat.pdf*. [98]

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [97]

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348. [107]

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995), "Adaptive Rejection Metropolis Sampling Within Gibbs Sampling," *Applied Statistics*, 44, 455–472. [107]

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman & Hall. [96]

Groisman, P. (2000), "Data Documentation for TD-3721: Gridded US Daily Precipitation and Snowfall Time Series," National Climatic Data Center, Asheville, NC. [111]

Gu, C. (2002), *Smoothing Spline ANOVA Models*, Berlin: Springer-Verlag. [96]

Härdle, W. K., and Scott, D. W. (1992), "Smoothing by Weighted Averaging of Rounded Points," *Computational Statistics*, 7, 97–128. [98]

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations*, Berlin: Springer-Verlag. [105,106]

Higdon, D., Swall, J., and Kern, J. (1999), "Non-Stationary Spatial Modeling," in *Bayesian Statistics* 6, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. Smith, Oxford: Clarendon Press, pp. 761–768. [98]

Hobert, J. P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models," *Journal of the American Statistical Association*, 91, 1461–1473. [104]

Holmes, C. C., and Mallick, B. K. (2001), "Bayesian Regression With Multivariate Linear Splines," *Journal of the Royal Statistical Society, Ser. B*, 63, 3–17. [98,109]

Kneib, T. (2006), "Mixed Model Based Inference in Structured Additive Regression," Ph.D. thesis, LMU München, Faculty of Mathematics, Computer Science and Statistics. [101,103]

Knorr-Held, L. (2003), "Some Remarks on Gaussian Markov Random Field Models for Disease Mapping," in *Highly Structured Stochastic Systems*, eds. N. H. P. Green and S. Richardson, Oxford: Oxford University Press, pp. 260–264. [97]

Knorr-Held, L., and Richardson, S. (2003), "A Hierarchical Model for Space-Time Surveillance Data on Meningococcal Disease Incidence," *Journal of the Royal Statistical Society, Ser. C*, 52, 169–183. [107]

Krivobokova, T., Crainiceanu, C. M., and Kauermann, G. (2008), "Fast Adaptive Penalized Splines," *Journal of Computational and Graphical Statistics*, 17, 1–20. [97,109]

Lang, S., and Brezger, A. (2004), "Bayesian P-Splines," *Journal of Computational and Graphical Statistics*, 13, 183–212. [97,109]

Lang, S., Fronk, E. M., and Fahrmeir, L. (2002), "Function Estimation With Locally Adaptive Dynamic Models," *Computational Statistics*, 17, 479–499. [97,103,114]

Mackay, D. J. C., and Takeuchi, R. (1998), "Interpolation Models With Multiple Hyperparameters," *Statistics and Computing*, 8, 15–23. [98]

Nychka, D. W. (2000), "Spatial-Process Estimates as Smoothers," in *Smoothing and Regression: Approaches, Computation, and Application*, ed. M. G. A. Schimek, New York: Wiley, pp. 393–424. [96]

Paciorek, C. J., and Schervish, M. J. (2004), "Nonstationary Covariance Functions for Gaussian Process Regression," in *Advances in Neural Information Processing Systems*, Vol. 16, eds. S. Thrun, L. Saul, and B. Schölkopf, Cambridge, MA: MIT Press. [98,109]

——— (2006), "Spatial Modelling Using a New Class of Nonstationary Covariance Functions," *Environmetrics*, 17, 483–506. [98]

Patra, M., and Karttunen, M. (2006), "Stencils With Isotropic Discretization Error for Differential Operators," *Numerical Methods for Partial Differential Equations*, 22, 936–953. [100]

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability*, Vol. 104, London: Chapman & Hall. [97,100,103,108]

Ruppert, D., and Carroll, R. J. (2000), "Spatially-Adaptive Penalties for Spline Fitting," *Australian & New Zealand Journal of Statistics*, 42, 205–223. [97]

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119. [98]

Schmidt, A. M., and O'Hagan, A. (2003), "Bayesian Inference for Non-Stationary Spatial Covariance Structure via Spatial Deformations," *Journal of the Royal Statistical Society, Ser. B*, 65, 743–758. [98]

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley. [98]

Speckman, P. L., and Sun, D. (2003), "Fully Bayesian Spline Smoothing and Intrinsic Autoregressive Priors," *Biometrika*, 90, 289–302. [97,104,113]

Sun, D., and Speckman, P. L. (2008), "Bayesian Hierarchical Linear Mixed Models for Additive Smoothing Splines," *Annals of the Institute of Statistical Mathematics*, 60, 499–517. [105]

Terzopoulos, D. (1988), "The Computation of Visible-Surface Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 417–438. [103]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia: SIAM. [96]

White, G. (2006), "Bayesian Semi-Parametric Spatial and Joint Spatio-Temporal Smoothing," Ph.D. thesis, University of Missouri–Columbia, Columbia, Missouri. [105]

Wood, S., Jiang, W., and Tanner, M. (2002), "Bayesian Mixture of Splines for Spatially Adaptive Nonparametric Regression," *Biometrika*, 89, 513–528. [98]

Yue, Y., Speckman, P. L., and Sun, D. (2008), "Fully Bayesian Adaptive Spline Smoothing," unpublished manuscript, University of Missouri–Columbia, Dept. of Statistics. [97,103-106,114]

Zellner, A. (1986), "On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, eds. P. K. Goel and A. Zellner, New York: Elsevier, pp. 233–243. [104]

Zellner, A., and Siow, A. (1980), "Posterior Odds Ratios for Selected Regression Hypotheses." in *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia (Spain)*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Valencia: Valencia University Press, pp. 585–603. [104]

Zhu, Z., and Wu, Y. (2010), "Estimation and Prediction of a Class of Convolution-Based Spatial Nonstationary Models for Large Spatial Data," *Journal of Computational and Graphical Statistics*, 19, 74–95. [98,113]