Class report on the paper "Nonstationary Spatial GMRFs" Yue and Speckman (2010)

Antonio Olivas-Martinez
Department of Biostatistics
University of Washington
Seattle, WA 98195
aolivas@uw.edu

Seth D Temple
Department of Statistics
University of Washington
Seattle, WA 98195
sdtemple@uw.edu

Abstract

Estimating smooth spatial processes from noisy observations is the focus of researchers from many fields. For surfaces with highly varying regions, many existing methods perform poorly. In this paper, the authors proposed a Bayesian method to overcome this issue. They built a second-order GMRF on a regular lattice by discretizing the penalty function of the thin-plate spline via second-order backward difference approximations of the Laplacian operator over the grid, with special care paid to the boundaries. These boundary corrections promoted an efficient implementation for (nonstationary) adaptive variance. In simulations and an applied example they demonstrated how the method improves surface inference in high varying regions. However, we showed that the method does fail when there are multiple regions of high variability clustered together. Finally, we offer critical remarks and suggested extensions for the Bayesian thin-plate spline (BATS).

1 Introduction

Yue and Speckman (2010) introduced a Bayesian method to estimate and forecast a smooth spatial process from noisy observations. Their approach is based on a discretized version of the thin-plate splines estimator. Consider the nonparametric regression model

$$y_i = f(u_i, v_i) + \epsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

where f is a smooth function we want to estimate, and $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \tau^{-1})$. The thin-plate spline estimator of f is the solution to the minimization problem

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

where λ is a smoothing parameter, and $J_2(f)$ is the roughness penalty for f on $\mathbb{R}^2$ defined by

$$J_2(f) = \iint_{\mathbb{R}^2} \left[\left(\frac{\partial^2 f}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial u \partial v} \right)^2 + \left(\frac{\partial^2 f}{\partial v^2} \right)^2 \right] du\, dv. \tag{3}$$

Another common approach to do spatial smoothing is kriging Cressie (2015). For large data sets, kriging is computationally intense. Moreover, in the nonstationary case, it may be hard to guess the trend. Interestingly, Nychka (2000) showed that kriging estimates are a type of spline provided that the basis functions and covariance matrix are correctly specified. Unfortunately, it is well known that smoothing splines perform badly when the function we want to estimate is highly varying (has peaks,

jumps or frequent curvature transitions) Eubank (1999). To overcome this challenge, methods that incorporate a spatially adaptive variance component have been proposed.

In the one-dimensional case, Yue et al. (2012) et al. proposed a Bayesian adaptive smoothing splines on a lattice. By using the discretized version of the spline estimator, they proposed a second-order intrinsic Gausian Markov random field (GMRF) prior with an adaptive variance component and a further first-order intrinsic GMRF for the variance function. They derived conditions on the prior distributions that guarantee the posterior is proper. The method proposed by Yue and Speckman (2010) is the two-dimensional extension of this method.

Brezger et al. (2007) et al. proposed another Bayesian model consisting on an adaptive GMRF with stochastic interaction weights in a space-varying coefficient model. However, as pointed out by Lang et al. (2002), this model fails for functions with high spatial variability. Modifications to fix this issue seem to be impractical or computationally infeasible as the model complexity grows. The method of Yue and Speckman (2010) can provide reliable inference in the case of extreme spatial variability and retains nice Markov properties for fast computation.

2 Bayesian adaptive thin-plate splines on a lattice

Consider the discretization of the regression model (1) over a regular lattice of $n = n_1 \times n_2$ points

$$y_{jkl} = f(\tilde{u}_j, \tilde{v}_k) + \epsilon_{jkl}, \quad \epsilon_{jkl} \overset{iid}{\sim} \mathcal{N}(0, \tau^{-1}),$$

for $j = 1, \ldots, n_1$, $k = 1, \ldots, n_2$, and $l = 1, \ldots, r_{jk}$. Coordinates $(\tilde{u}_j, \tilde{v}_k)$ correspond to points on the lattice, and r_{jk} is the number of observations at location $(\tilde{u}_j, \tilde{v}_k)$. Let $N = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} r_{jk}$ be the total number of observations and $x_1, \ldots, x_n$ be the locations ordered by columns. The vectorized version of the discretized regression model is:

$$\mathbf{y} = \mathbf{D}\mathbf{z} + \epsilon, \quad \epsilon \sim \mathcal{N}_N(\mathbf{0}, \tau^{-1}\mathbf{I}_N), \tag{4}$$

where $\mathbf{y} = (y_{x_1 1}, \ldots, y_{x_n r_{x_n}})'$, $\mathbf{z} = (f(x_1), \ldots, f(x_n))'$, $\epsilon = (\epsilon_{x_1 1}, \ldots, \epsilon_{x_n r_{x_n}})'$, and $D = [d_{ij}]$ is the incidence matrix: $d_{ij} = 1$ if observation i in $\mathbf{y}$ is at location x_j.

2.1 Discretized roughness penalty

The proposed method performs Bayesian inference on $\mathbf{z}$, so a prior on $\mathbf{z}$ is required. To get such a prior, the authors built a bivariate intrinsic GMRF based on an approximation of the Laplacian operator. Their intrinsic GMRF is similar to that seen in (Rue and Held, 2005, pp. 114-116) but with some modifications to the lattice boundaries.

For functions $f : \mathbb{R}^2 \to \mathbb{R}$ with integrable derivatives of order up to 3 and that vanish at infinity (i.e. $f \in \mathcal{C}_0^4(\mathbb{R}^2)$), integrating by parts, we obtain that the rough penalty $J_2(f)$ defined in (3) can be written as

$$J_2(f) = \iint \left[\Delta f(u, v)\right]^2 du\, dv, \tag{5}$$

where $\Delta = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}$ is the bivariate Laplacian differential operator. For h small, the second order derivatives can be approximated by

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

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

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

Then, the roughness penalty (5) can be approximated by

$$\frac{1}{h^4} \sum_{j=2}^{n_1 - 1} \sum_{k=2}^{n_2 - 1} \left[\nabla^2 f(\tilde{u}_j, \tilde{v}_k)\right]^2, \tag{6}$$

where

$$\nabla^2 f(\tilde{u}_j, \tilde{v}_k) = \left(\nabla^2_{(1,0)} + \nabla^2_{(0,1)}\right) f(\tilde{u}_{j+1}, \tilde{v}_{k+1}).$$

Noting that the expression in (6) is a quadratic form on $\mathbf{z}$, it can be written as $\mathbf{z}'\mathbf{A}_0\mathbf{z}$ with $\mathbf{A}_0$ a positive semidefinite matrix. Then, the discretized version of the thin-spline estimator in (2) is

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[(\mathbf{y} - \mathbf{D}\mathbf{z})'(\mathbf{y} - \mathbf{D}\mathbf{z}) + \frac{\mathbf{\Lambda}}{h^4}\mathbf{z}'\mathbf{A}_0\mathbf{z} \right]. \tag{7}$$

Assuming a hierarchical model where $\mathbf{y}$ is (conditionally) multivariate normal, the optimization problem (7) suggests that the prior on $\mathbf{z}$ should have the form

$$[\mathbf{z}|\delta] \propto \delta^{(n-m)/2} |\mathbf{A}_0|_+^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}_0\mathbf{z} \right). \tag{8}$$

so that the posterior on $\mathbf{z}$ has $(\mathbf{y} - \mathbf{D}\mathbf{z})'(\mathbf{y} - \mathbf{D}\mathbf{z}) + \frac{\mathbf{\Lambda}}{h^4}\mathbf{z}'\mathbf{A}_0\mathbf{z}$ in its (multivariate normal) exponent.

2.2 Boundary correction

Expression (6) only sums from $j = 2$ to $j = n_1 - 1$ and from $k = 2$ to $k = n_2 - 1$ because second order backward differences are not available at the lattice boundaries. Moreover, this matrix $\mathbf{A}_0$, hereafter referred to as a structure matrix, may have a null space up to $2(n_1 + n_2 - 2)$ dimensions. Whereas RW1 and RW2 models on regular locations have structure matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ such that the distribution of $\mathbf{z}$ is invariant to the addition of a constant or a linear function of the indices, the null space of structure matrix $\mathbf{A}_0$ does not lend as convenient an interpretation.

To address these issues, Yue and Speckman (2010) put forth additional differential operators to refine the discrete approximation at the corners and edges and arrived at a structure matrix $\tilde{\mathbf{A}}$ with more interpretable rank-deficiency. For the edges, depending on the edge, a second order backward difference is combined with a first order backward difference $(+\frac{\partial^2}{\partial \tilde{u}^2} \pm \frac{\partial}{\partial \tilde{v}})$ or $(\pm\frac{\partial}{\partial \tilde{u}} + \frac{\partial^2}{\partial \tilde{v}^2})$; for the corners, first order backward differences are proposed $(\pm\frac{\partial}{\partial \tilde{u}} \pm \frac{\partial}{\partial \tilde{v}})$. These result in eight new differential operators ∇_i for $i = 1, \ldots, 8$, forming the improved discrete approximation

$$\begin{aligned}
&\frac{1}{h^4}\Bigg[\sum_{j=2}^{n_1-1}\sum_{k=2}^{n_2-1} \{\nabla_0^2 f(\tilde{u}_j, \tilde{v}_k)\}^2 \\
&+ \underbrace{\{\nabla_1 f(\tilde{u}_1, \tilde{v}_1)\}^2 + \{\nabla_1 f(\tilde{u}_1, \tilde{v}_1)\}^2 + \{\nabla_2 f(\tilde{u}_{n_1}, \tilde{v}_1)\}^2 + \{\nabla_3 f(\tilde{u}_1, \tilde{v}_{n_2})\}^2 + \{\nabla_4 f(\tilde{u}_{n_1}, \tilde{v}_{n_2})\}^2}_{\text{corners}} \\
&+ \underbrace{\sum_{j=2}^{n_1-1} \{\nabla_5 f(\tilde{u}_j, \tilde{v}_1)\}^2 + \{\nabla_6 f(\tilde{u}_j, \tilde{v}_{n_2})\}^2 + \sum_{k=2}^{n_2-1} \{\nabla_7 f(\tilde{u}_1, \tilde{v}_k)\}^2 + \{\nabla_8 f(\tilde{u}_{n_1}, \tilde{v}_k)\}^2}_{\text{edges}} \Bigg]
\end{aligned} \tag{9}$$

As before, (9) is quadratic in $\mathbf{z}$ and can be written $\mathbf{z}'\tilde{\mathbf{A}}\mathbf{z}$ with respect to the structure matrix $\tilde{\mathbf{A}}$. Additionally, the marginal terms in (9) are all squared, suggesting the simple factorization $\tilde{\mathbf{A}} = \tilde{\mathbf{B}}'\tilde{\mathbf{B}}$. The boundary correction is most important in that the null space of $\tilde{\mathbf{A}}$ is one-dimensional and spanned by the constant vector. That is, GMRF priors with this structure matrix $\tilde{\mathbf{A}}$ have the same desirable invariance property as the RW1 model.

Lastly, Yue and Speckman (2010) remove the first row of $\tilde{\mathbf{B}}$ to get $\mathbf{A} = \mathbf{B}'\mathbf{B}$ where $\mathbf{B}$ has full rank $n - 1$. The full rank property of the modified structure matrix accommodates an efficient nonstationary extension with adaptive variance. The prior on $\mathbf{z}$ is now posed as

$$[\mathbf{z}|\delta] \propto \delta^{(n-1)/2} |\mathbf{A}|_+^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}\mathbf{z} \right). \tag{10}$$

The structure matrix $\mathbf{A}$ for this GMRF prior has the dependence structures with respect to the regular lattice shown in Figure 1. These are encoded in the matrix representation to promote an efficient banded Cholesky decomposition with bandwidth at most $2n_1$.

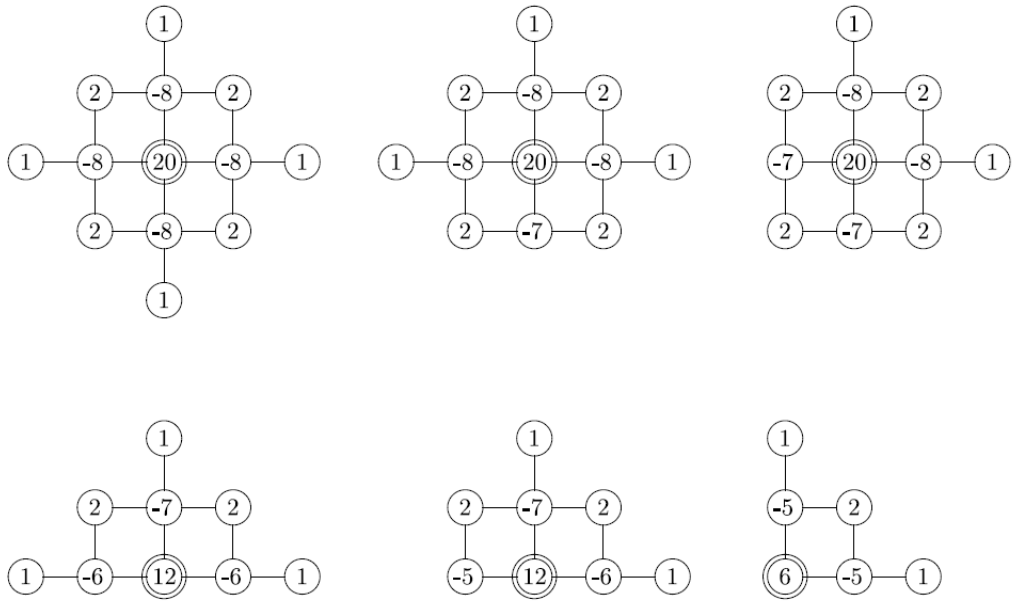

Figure 1: (Top left) Interior dependence, (top middle) near edge dependence, (top right) near corner dependence, (bottom left) edge dependence, (bottom middle) near corner, on edge dependence, (bottom right) corner dependence.

2.3 Adaptive variance

An alternative representation for the prior on $\mathbf{z}$ is

$$\mathbf{Bz} \sim \mathcal{N}_{n-1}(\mathbf{0}, \delta^{-1}\mathbf{I}_{n-1}),$$

which means that the differences over the grid, except for the one in the upper left corner, are i.i.d. Gaussian with mean zero and variance δ^{-1}. To do the adaptive extension, the authors replaced δ with $(n-1)$ locally varying precision δ_{jk}'s. These precision parameters should be large in flat areas and small in highly changing regions. Specifically, the authors expressed $\delta_{jk} = \delta e^{\gamma_{jk}}$ with positive scale parameter δ and the linear constrain $\sum \gamma_{jk} = 0$ for identifiability. Next, they assumed a first-order IGMRF prior on a regular lattice for $\gamma = (\gamma_{jk})'$. With this extension, the nonstationary GMRF prior for the adaptive model is

$$[\mathbf{z}|\delta, \gamma] \propto \delta^{(n-1)/2} |\mathbf{A}_\gamma|_+^{1/2} \exp\left(-\frac{\delta}{2}\mathbf{z}'\mathbf{A}_\gamma z\right), \tag{11}$$

$$[\gamma|\eta] \propto \eta^{(n-2)/2} \exp\left(-\frac{\eta}{2}\gamma'\mathbf{M}\gamma\right) I(\mathbf{1}'\gamma = 0). \tag{12}$$

where the adaptive structure matrix $\mathbf{A}_\gamma = \mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B}$ depends on $\mathbf{\Lambda}_\gamma = \text{diag}(e^{\gamma_{jk}})$ and the null space of $\mathbf{M}$ is spanned by the constant vector $\mathbf{1}$. The proposed nonstationary GMRF prior (11 and 12) has appealing properties for Bayesian inference:

1. It improves surface smoothing.
2. A Gibbs sampler is easy to implement and sparsity promotes efficient MCMC simulation (see Section 3).
3. $|\mathbf{A}_\gamma|_+ = |\mathbf{B}\mathbf{B}'|$ is constant, so that $|\mathbf{A}_\gamma|_+^{1/2}$ is computed once (see Lemma 2.1).
4. The modified GMRF prior (11) is second order in the interior, yielding more smoothness and flexibility for estimation.

Lemma 2.1. *For the adaptive structure matrix* $\mathbf{A}_\gamma = \mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B}$*, it holds that* $|\mathbf{A}_\gamma|_+ = |\mathbf{B}\mathbf{B}'|$.

Proof. Let $\mathbf{B} = \mathbf{PDQ}'$ be the singular value decomposition of $\mathbf{B}$ where $\mathbf{D} = [\ \tilde{\mathbf{D}} \quad \mathbf{0}\ ]$ and $\tilde{\mathbf{D}}$ is a $(n-1) \times (n-1)$ diagonal matrix whose diagonal entries are positive. Let $\mathbf{Q} = [\ \tilde{\mathbf{Q}} \quad q_n\ ]$ where $\tilde{\mathbf{Q}}$ is a $n \times (n-1)$ matrix, then

$$\mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B} = \mathbf{QD}'\mathbf{P}'\mathbf{\Lambda}_\gamma\mathbf{PDQ}' = \tilde{\mathbf{Q}}\tilde{\mathbf{D}}\mathbf{P}'\mathbf{\Lambda}_\gamma\mathbf{P}\tilde{\mathbf{D}}\tilde{\mathbf{Q}}'.$$

which implies that

$$\begin{aligned}
|\mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B} - \lambda\mathbf{I}_n| &= |\mathbf{Q}'(\mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B} - \lambda\mathbf{I}_n)\mathbf{Q}| \\
&= \left| \begin{bmatrix} \mathbf{I}_{n-1} \\ \mathbf{0}' \end{bmatrix} \tilde{\mathbf{D}}\mathbf{P}'\mathbf{\Lambda}_\gamma\mathbf{P}\tilde{\mathbf{D}} \begin{bmatrix} \mathbf{I}_{n-1} & 0 \end{bmatrix} - \lambda\mathbf{I}_n \right| \\
&= \left| \begin{bmatrix} \tilde{\mathbf{D}}\mathbf{P}'\mathbf{\Lambda}_\gamma\mathbf{P}\tilde{\mathbf{D}} & 0 \\ \mathbf{0}' & 0 \end{bmatrix} - \lambda\mathbf{I}_n \right|.
\end{aligned}$$

Since $\mathbf{P}$ is positive definite, it is clear that the non-zero eigenvalues of $\mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B}$ are given by the solutions to the equation $|\tilde{\mathbf{D}}\mathbf{P}'\mathbf{\Lambda}_\gamma\mathbf{P}\tilde{\mathbf{D}} - \lambda\mathbf{I}_{n-1}| = 0$. Noting that $|\mathbf{\Lambda}_\gamma| = 1$, that $\mathbf{P}'\mathbf{P} = \mathbf{I}_{n-1}$ and that the product of the eigenvalues of a matrix equals its determinant, we get $|\mathbf{B}'\mathbf{\Lambda}_\gamma\mathbf{B}|_+ = |\tilde{\mathbf{D}}|^2$. But we also have

$$|\mathbf{BB}'| = |\mathbf{PDQ}'\mathbf{QD}'\mathbf{P}'| = |\mathbf{P}\tilde{\mathbf{D}}\tilde{\mathbf{D}}\mathbf{P}'| = |\tilde{\mathbf{D}}|^2 \ \square$$

$\square$

3 Implementation

Bayesian thin-plate splining for spatial smoothing as presented in Yue and Speckman (2010) involves posterior inference from MCMC sampling of the following hierarchical model

$$\begin{aligned}
\mathbf{y} \mid \tau, \mathbf{D}, \mathbf{z} &\sim \mathcal{N}_N(\mathbf{Dz}, \tau\mathbf{I}_N) \\
\mathbf{z} \mid \delta, \boldsymbol{\gamma} &\propto (\tau\xi_1)^{(n-1)/2} |\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2} \exp(-\delta/2 \cdot \mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}) \\
\boldsymbol{\gamma} \mid \eta &\propto \eta^{(n-2)/2} \exp(-\eta/2\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}) \\
\tau &\propto \tau^{-1} \\
\delta &\sim \text{some hyperprior} \\
\eta &\sim \text{some hyperprior}
\end{aligned}$$

This hierarchical model can be sampled in a Gibbs way because the posterior distributions by and large have nice forms. There are two challenges in sampling from this hierarchical model: (1) some of the steps in the Gibbs sampling procedure require specialized algorithms, and (2) the hyperpriors must be carefully specified so as to result in a posterior distribution for τ that is proper. The authors reparameterized this model with the one-to-one correspondence $(\tau, \delta, \eta) \to (\tau, \xi_1, \xi_2)$ where $\xi_1 = \delta\tau^{-1}$ and $\xi_2 = \eta\delta^{-1}$ and then specified hyperpriors for ξ_1 and ξ_2. Namely, $\xi_1 \sim$ Pareto(c) and $\xi_2 \sim$ Inverse-Gamma(a, b) independent.

3.1 Gibbs sampler

Given the hierarchical model, posterior distributions can be derived for the model parameters $\mathbf{z}, \boldsymbol{\gamma}, \tau$ and the hyperpriors ξ_1, ξ_2. For Gibbs sampling, these derivations involve writing at conditional densities, ignoring terms not attached to a given parameter using $\propto$, and simplifying algebra. Then, the full conditional distributions are

1. $[\mathbf{z} \mid \mathbf{y}, \mathbf{D}, \tau, \xi_1, \boldsymbol{\gamma}] \sim \mathcal{N}_n((\mathbf{D}'\mathbf{D} + \xi_1\mathbf{A}_{\boldsymbol{\gamma}})^{-1}, \mathbf{D}'\mathbf{y}, (\mathbf{D}'\mathbf{D} + \xi_1\mathbf{A}_{\boldsymbol{\gamma}})^{-1})$
2. $[\boldsymbol{\gamma} \mid \mathbf{z}, \xi_1, \xi_2] \propto \exp(-\tau\xi_1/2 \cdot \mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z} - \tau\xi_1\xi_2/2 \cdot \boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma}) I(\mathbf{1}'\boldsymbol{\gamma} = 0)$
3. $[\tau \mid \mathbf{y}, \mathbf{z}, \boldsymbol{\gamma}, \xi_1, \xi_2] \sim$ Gamma$(n + (N-3)/2, 0.5 \cdot ||y - \mathbf{Dz}||_2^2 + \xi_1/2 \cdot \mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z} + \xi_1\xi_2/2 \cdot \boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma})$
4. $[\xi_1 \mid \mathbf{z}, \tau, \xi_1] \sim$ Gamma$((n-1)/2, \tau/2 \cdot \mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z} + \tau\xi_2/2 \cdot \boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma} + \theta)$
5. $[\xi_2 \mid \boldsymbol{\gamma}, \xi_1] \propto \xi_2^{(n-2)/2-a-1} \exp(-\tau\xi_1\xi_2/2 \cdot \boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma} - b/\xi_2)]$

6. $[\theta \mid \xi_1] \sim \text{Gamma}(2, \xi_1 + c)$

where the Pareto prior for ξ_1 is represented as a scale mixture of exponential random variables through the latent hierarchy in θ. Step 1 may be sampled in a block and is efficient because the structure matrix $\mathbf{A}_{\boldsymbol{\gamma}}$ was constructed so that $\mathbf{D}'\mathbf{D} + \xi_1 \mathbf{A}_{\boldsymbol{\gamma}}$ to enable a banded Cholesky decomposition with low bandwidth.

Steps 2 and 5 require specialized algorithms to sample the full conditional distributions. Yue and Speckman (2010) note that the function expression in Step 5 is log concave and is therefore appropriate for adaptive rejection Metropolis sampling (ARMS) (Gilks and Wild, 1992). ARMS starts with a proposal density that envelopes the target density and iteratively modifies the proposal density during sampling so that samples become accepted more often over time. This technique is especially tuned to support log concave target densities that do not have explicit conjugate forms. Naively, Step 2 involves sampling from an intrinsic GMRF with a hard constraint for which Algorithm 6 in Lecture 7 is appropriate (Roulet, STAT 517, 2022). However, the identifiability constraint $\mathbf{1}'\boldsymbol{\gamma} = 0$ induces certain relationships among the γ_{jk} model parameters, making it more difficult to explore the posterior distribution. Naive implementations of block sampling for $\boldsymbol{\gamma}$ result in block sums that never change. Yue and Speckman (2010) suggest block sampling by rows in one iteration and then by columns in another iteration to break up these dependencies and better explore the posterior.

3.2 Equivalent degrees of freedom

The global parameter τ controls the measurement error and is commonly given a noninformative Jeffrey's prior in the Bayesian setting. Various literature in spatial statistics indicates that such an improper prior $\tau \propto \tau^{-1}$ leads to an improper posterior. Yue and Speckman (2010) opine that Bayesian inference in the smoothing problem is impossible without some informative priors, which they implement through proper priors on ξ_1 and ξ_2.

Theorem 3.1. *For the nonparametric thin-plate splite model with the Jeffrey's prior $\tau \propto \tau^{-1}$, the following sufficient conditions ensure a proper joint posterior $(\mathbf{z}, \tau, \xi_1, \xi_2, \boldsymbol{\gamma})$:*

1. *(Repeated measurements) $[\xi_1, \xi_2]$ is proper*
2. *(Complete, non-repeated measurements) $[\xi_1, \xi_2]$ is proper and $\mathbb{E}[\xi_2^{-(N-1)/2}]$ is finite*
3. *(Incomplete, non-repeated measurements) $[\xi_1, \xi_2]$ is proper and $\mathbb{E}[(\xi_1\xi_2)^{-(N-1)/2}]$ is finite*

Independent $\xi_1 \sim \text{Pareto}(c)$ and $\xi_2 \sim \text{Inverse-Gamma}(a, b)$ are given as proper priors that satisfy these sufficient conditions. These choices were not motivated or discussed in the paper. No other priors were discussed or explored.

Second, Yue and Speckman (2010) recommend a new procedure to specify hyperpriors in Bayesian spatial smoothing. Their suggestion is based on "equivalent degrees of freedom" as seen in (Hastie et al., 2001, Chapter 5). In linear models, linear predictions $\mathbf{P}\mathbf{y}$ depend on projection $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$. The trace of the projection matrix $\mathbf{P}$ turns out to be equal to the rank of the design matrix $\mathbf{X}$, where rank is the number of linearly independent covariates. Analogously, in smoothing problems, a smoother matrix $\mathbf{S}$ is applied to data $\mathbf{y}$, and the trace of the smoother matrix $\mathbf{S}$ is referred to as the "equivalent degrees of freedom". In both settings, the degrees of freedom serves as a proxy for model complexity. For the nonstationary spatial GMRF prior, there is spatial smoothing in $\mathbf{z}$ and spatial smoothing in $\boldsymbol{\gamma}$. Fixing the hyperparameter a in the Inverse-Gamma hyperprior, the authors determine b and c such that the trace of the associated smoother matrices equals prespecified "equivalent degrees of freedom". That is, the practitioner must state *a priori* how complex they believe their spatial model is. Alternatively, Yue and Speckman (2010) speculate that $\text{df}_1 \approx 50$ and $\text{df}_2 \approx \min(100, n/3)$ may be okay rules for hyperprior specification. We find these aspects of their paper to be lacking, especially given that the priors exert strong influence on the posterior inference and that trace plots for ξ_1, ξ_2 can demonstrate poor convergence (Yue and Speckman, 2010, Figure 5).

4 Experimental Results

In this section, we present two simulation studies and provide a summary of the applied example provided in the paper. These experimental results demonstrate that a Bayesian thin-plate spline model with the proposed nonstationary spatial GMRF prior can accurately model drastic changes to spatial dependence and variability.

4.1 Simulation study 1. Function with two uneven modes

This simulation study was provided in the paper. We chose to present only this simulation because it involves a surface with a highly changing region (a mode with a sharp peak). The aim is to compare the proposed method (Bayesian adaptive thin-plate splines, BATS) against the same method without the adaptive component (nonadaptive BTS) to estimate the following mean surface function

$$f(u,v) = 2\exp\left\{-\frac{1}{0.4}\left[(u-2)^2+(v-2)^2\right]\right\} + \exp\left\{-\frac{1}{3}(u^2+v^2)\right\}, \quad u,v \in (-5,5).$$

The authors used a sample size $n = 900$ with $n_1 = n_2 = 30$, $\tau^{-1} = 0.1^2$ and 250 simulations for the function. They also chose $a = 0.5$ and $b = 0.001$ which correspond to approximately 200 median prior degrees of freedom for the hyperparameter ξ_2, and $c = 8$ that corresponds to about 50 median prior degrees of freedom for the hyperparameter ξ_1. Results are shown in Figure 2. As can be seen, the nonadaptive method could not capture the sharp peak whereas the adaptive method slightly undersmoothed the small bump. Interestingly, the estimated variance only had a meaningful variation around the sharp peak.

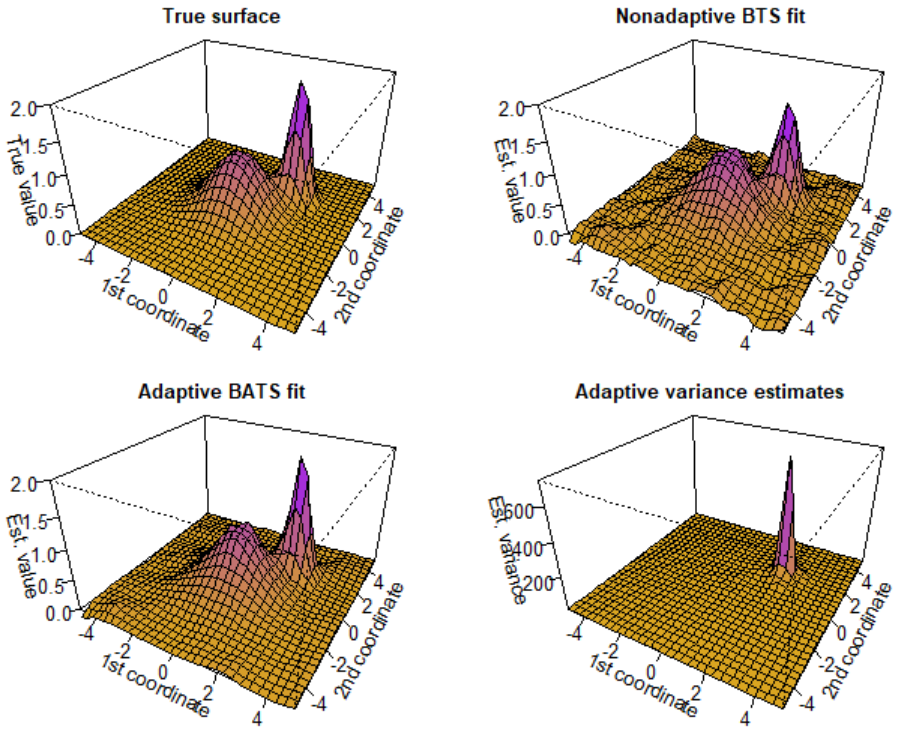

Figure 2: Bimodal simulated example: grid of 30×30, $\tau^{-1} = 0.1^2$, and $f(u,v) = 2\exp\left\{-\frac{1}{0.4}\left[(u-2)^2+(v-2)^2\right]\right\} + \exp\left\{-\frac{1}{3}(u^2+v^2)\right\}$ for $u,v \in (-5,5)$.

4.2 Simulation study 2. Multiple modes

Next, we aim to estimate a mean regression function with four sharp peaks for local modes. We explored what happened when the modes got closer to each other as well. We used the same grid and

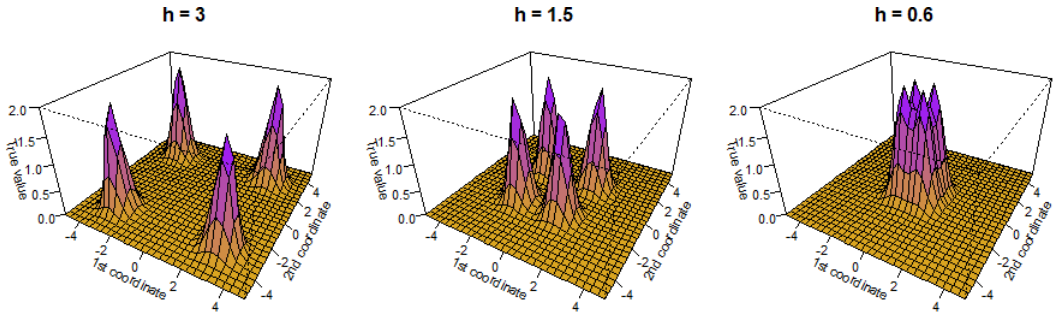

Figure 3: True surface: left ($h = 3$), middle ($h = 1.5$), right ($h = 0.6$), grid of 30×30 and $\tau^{-1} = 0.1^2$

the same precision for the errors as in the previous study. The mean regression surfaces are then

$$f(u,v) = 2\exp\left\{-\frac{1}{0.4}\sum_{\kappa_1,\kappa_2\in\{-1,1\}}\left[(u-\kappa_1 h)^2 + (v-\kappa_2 h)^2\right]\right\}, \quad u,v \in (-5,5),$$

for $h = 3, 1.5$ and 0.6. The true surfaces are displayed in Figure 3.

The BATS fit and the estimated adaptive variance for $h = 3, 1.5$ and 0.6 are shown in Figure 4. As can be observed, we found estimation of the modes to be more challenging the closer they got. This is consistent with what we see in the estimated adaptive variances: when the modes are close, the variances exhibit noisy behavior around the modal locations.

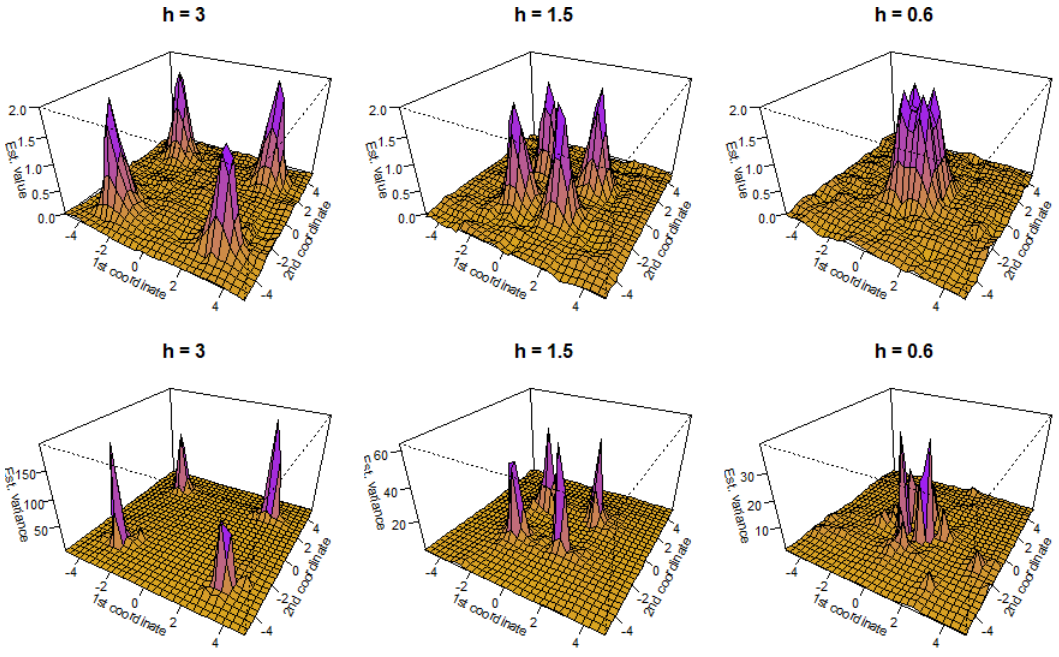

Figure 4: Case: left ($h = 3$), middle ($h = 1.5$), right ($h = 0.75$), top (Adaptive BATS fit), bottom (Estimated variance)

4.3 Application to U.S. rainfall data

Unfortunately, the authors provided a dataset with the response variable rainfall but no locations, so we could not replicate their example. Here we present a mere summary of their results. The

data consisted of 2082 annual precipitation measurements for year 1960 irregularly spaced over a rectangular region. Overall, they observed higher rainfall in the south than in the north. To implement the proposed method, they binned the data to a 60×60 regular lattice, resulting in 57% of the grid points with no data values. They observed that the BATS model handled the missing observations well. They also compared the BATS fit against an nonadaptive thin-plate spline method (TPS). The inferred surface from BATS was more peaked in mountainous regions than the inferred surface from TPS. Otherwise, both methods inferred smooth surfaces in the rest of the regions. In conclusion, BATS identifies high variability in spatial processes in environmentally extreme regions, e.g., mountains, whereas other methods like TPS oversmooth.

5 Discussion and conclusions

In review, Yue and Speckman (2010) posit a two level nonstationary spatial GMRF prior for Bayesian hierarchical modeling of the TPS problem that has (1) a flexible second-order neighborhood structure in the lattice interior and (2) adaptive variance for applied setting where environmental landscapes change abruptly. Their contribution to spatial statistics is mainly theoretical and methodological and less so application-focused. For instance, their supplemental code is written in Fortran, a low-level programming language already out of fashion in 2010, and difficult to modify for average and even sophisticated users. An application-focused implementation on this paper would be to implement its methodology in a high-level programming language with a user-friendly interface and additional utitilities for parallel processing and MCMC chains. Second, the provided code only runs the second case with complete, non-repeated measurements, so the repeated measurements and missing data scenarios would need to be implemented to reach a larger audience. Third, as mentioned in their summary, an omnibus model with trend effects

$$\begin{aligned} \mathbf{y} \mid \mathbf{D}, \mathbf{z}, \mathbf{X}, \boldsymbol{\beta} &\sim \mathcal{N}_N(\mathbf{D}\mathbf{z} + \mathbf{X}\boldsymbol{\beta}, \tau \mathbf{I}_N) \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ \mathbf{z} \mid \delta, \boldsymbol{\gamma} &\propto (\tau\xi_1)^{(n-1)/2} |\mathbf{A}_{\boldsymbol{\gamma}}|_+^{1/2} \exp(-\delta/2 \cdot \mathbf{z}'\mathbf{A}_{\boldsymbol{\gamma}}\mathbf{z}) \end{aligned}$$

could be implemented by deriving the full conditionals in the Gibbs sampler as an extension. We emphasize this extension as detrending is standard practice in spatial analyses. Few responses in spatial modeling are separate from environmental covariates. We believe it would be interesting to reanalyze the applied rainfall example but with the covariates longitude and latitude. Finally, BATS model of Yue and Speckman (2010) requires input data on a regular lattice. We tried to apply the BATS model to lead (Pb) measurements in Lake Leman, Switzerland. On one hand, the lake is arced, not rectangular, so the only part of the lake can be modeled in the BATS framework. On the other hand, samples were collected in a hexagonal pattern, so we had to bin measurements into a regular lattice based on an average over nearby points. While Yue and Speckman (2010) cite literature on nonparametric regression and density estimation to write off the implications of binning, we reserve judgment until a comprehensive simulation study is conducted. For areal data, e.g. public health units in Ontario, Canada, states in the USA, etc., Yue and Wang (2014) propose a GMRF framework for irregular, possibly non-rectangular lattices based on adjacency matrices. Extending the nonstationary spatial GMRF prior method for irregular lattices in higher resolution spatial datasets would be a significant contribution to the spatial statistics field. These and other improvements to the methodology could better popularize these excellent theoretical advances.

References

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: Series C (Applied Statistics)*, 56(3):327–345.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *J. R. Stat. Soc. Ser. C Appl. Stat.*, 41(2):337.

Hastie, T., Friedman, J., and Tibshirani, R. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY.

Lang, S., Pronk, E.-M., and Fahrmeir, L. (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics*, 17(4):479–499.

Nychka, D. W. (2000). Spatial-process estimates as smoothers. *Smoothing and regression: approaches, computation, and application*, 329:393.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC.

Yue, Y. and Speckman, P. L. (2010). Nonstationary spatial gaussian markov random fields. *J. Comput. Graph. Stat.*, 19(1):96–116.

Yue, Y. R., Speckman, P. L., and Sun, D. (2012). Priors for bayesian adaptive spline smoothing. *Annals of the Institute of Statistical Mathematics*, 64(3):577–613.

Yue, Y. R. and Wang, X.-F. (2014). Spatial gaussian markov random fields: Modelling, applications and efficient computations.