Locally-Adaptive Spatial Smoothing with Shrinkage-Prior Markov Random Fields

James Faulkner[1,2] and Vladimir Minin[3,4]

1.*Quantitative Ecology and Resource Management, University of Washington;* 2.*National Oceanic and Atmospheric Administration, Seattle, WA;* 3.*Departments of Statistics and Biology, University of Washington;* 4.*Department of Statistics, UC Irvine*

Motivation

Suppose we have spatially referenced data and want to make inference about the underlying spatial process. The process has breakpoints, sharp local features, or varying smoothness, which are all difficult features to capture with standard methods. We want a fully Bayesian method that can accurately estimate the surface, yet is easy to understand and straight forward to implement.

Current Bayesian methods

Some methods developed to deal with such situations include:

- Non-stationary Gaussian processes (*e.g.*, Paciorek and Schervish 2006)
- Non-Gaussian processes (*e.g.*, Bolin 2014)
- Adaptive Gaussian Markov random fields (GMRF; *e.g.* Yue et al. 2010, 2014)

We propose a type of adaptive GMRF that uses shrinkage priors to provide a balance of local sensitivity global smoothing.

Basic GMRF smoothing prior

Assume there is a process in continuous space that follows an unknown function $f(\boldsymbol{s})$, where $\boldsymbol{s} \in \mathbb{R}^2$. Let $\theta_i = \tilde{f}(\boldsymbol{a}_i)$ be the expected value of the surface over some discrete areal unit $\boldsymbol{a}_i$.Then a simple kth-order GMRF prior for $\boldsymbol{\theta}$ is induced by letting:

$$\Delta^k\theta_j \sim \mathrm{N}(0, \gamma^2)$$

where $\Delta^k\theta_j$ is a kth-order spatial difference operator. The resulting joint distribution for $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} \mid \boldsymbol{\mu}, \mathbf{Q} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$$

The precision matrix $\mathbf{Q}$ for the first-order ($k = 1$) model is given by,

$$Q_{ij} = \frac{1}{\gamma^2}\begin{cases} w_{i+} & j = i \\ -w_{ij} & j \neq i \\ 0 & \text{otherwise.} \end{cases}$$

Let d_{ij} be either the Euclidean distance between units i and j or a simple binary adjacency indicator. Then for neighbors of unit i,

$$w_{ij} = \frac{1}{d_{ij}}$$

and

$$w_{i+} = \sum_{j \sim i} \frac{1}{d_{ij}}.$$

To make $\mathbf{Q}$ positive-definite and make $p(\boldsymbol{\theta})$ proper, we set

$$Q_{11} = 1/\omega^2 + w_{1+},$$

where ω is a scale parameter related to the marginal variance of the θ's. We let the global scale parameter γ follow a half-Cauchy distribution:

$$\gamma \sim \mathrm{C}^+(0, \zeta)$$

Adaptive SPMRF smoothing prior

We can allow locally-adaptive behavior and increase smoothing properties by putting a **shrinkage prior** on $\Delta^k\theta_i$:

$$\Delta^k\theta_i \sim \text{Horseshoe}(0, \gamma)$$

$$\gamma \sim \mathrm{C}^+(0, \zeta)$$

where γ is the global smoothing parameter. The result is non-Gaussian Markov random field for $\boldsymbol{\theta}$, which we call a shrinkage-prior Markov random field (SPMRF), where in this case the shrinkage prior is the horseshoe distribution.

Adaptive SPMRF continued

There is no closed form for the marginal joint distribution of $\boldsymbol{\theta}$ for the horseshoe SPMRF, but we can introduce a set of latent variables $\boldsymbol{\tau}$ that allow a hierarchical representation where the joint distribution of $\boldsymbol{\theta}$ is normal conditional on $\boldsymbol{\tau}$:

$$\boldsymbol{\theta} \mid \mathbf{Q}(\boldsymbol{\tau}) \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{Q}(\boldsymbol{\tau})^{-1})$$

$$\tau_{ij} \mid \gamma \sim \mathrm{C}^+(0, \gamma)$$

$$\gamma \sim \mathrm{C}^+(0, \zeta)$$

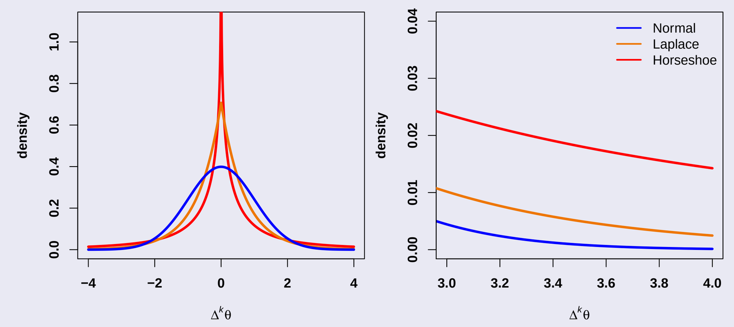
For the first-order model the τ_{ij} can be seen as scale parameters for the distributions of the pairwise differences $\theta_i - \theta_j$. Here the precision matrix is specified as in the GMRF but without γ and now

$$w_{ij} = \frac{1}{d_{ij}\tau_{ij}^2}$$

Shrinkage priors

A good shrinkage prior will:

- Shrink weak signals: *high mass near zero*
- Let strong signals through: *long tails*



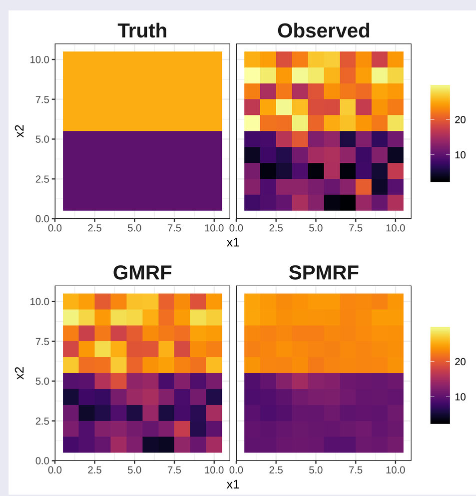
Posterior inference

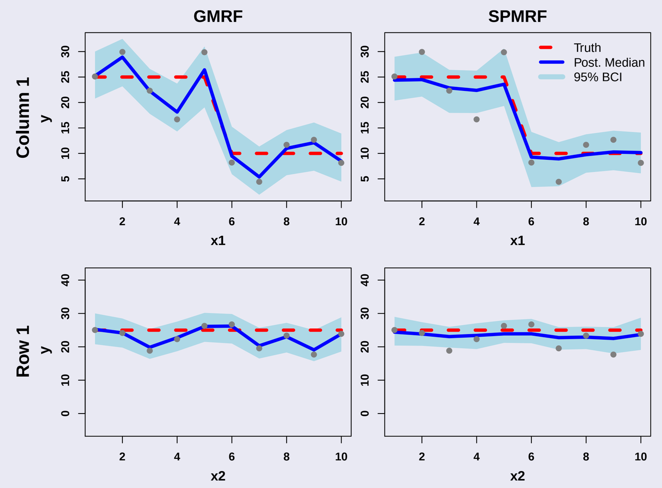
We use Hamiltonian Monte Carlo (HMC) for posterior inference. HMC uses Hamiltonian dynamics to generate joint parameter proposals. This proposal mechanism results in improved mixing and high acceptance rates, but the gradient calculations can be costly for some models. The sparse precision matrices allow the use of sparse matrix operations to reduce computation costs.

Simulated spatial processes

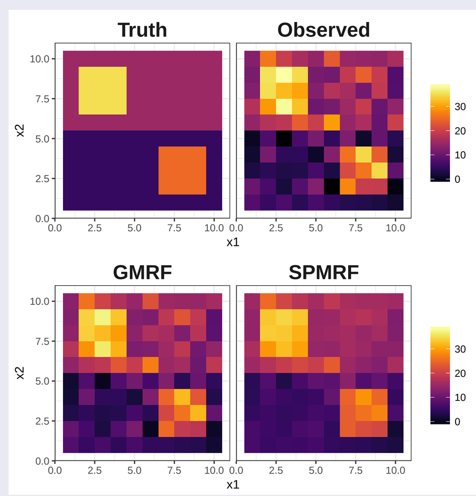
We investigate two scenarios, each on a 10 x 10 uniform grid with a single observation per grid cell. Observations are conditionally independent where $y_i \mid \boldsymbol{\theta} \sim \mathrm{N}(\theta_i, \sigma^2)$ and $\sigma^2 = 4$ in each scenario. We only investigate results for a single realization for each process.

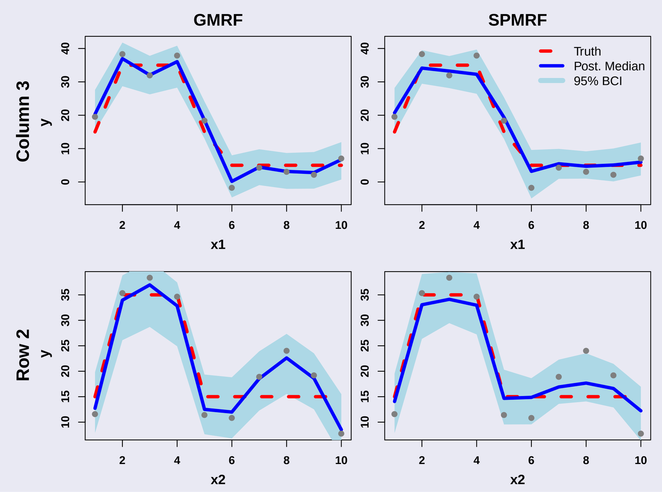
Scenario 1:





Scenario 2:

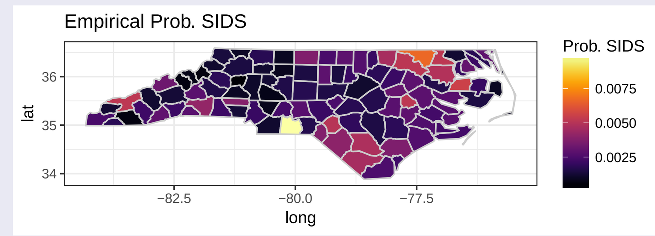




Data example: SIDS

Data are counts of incidence of sudden infant death syndrome (SIDS) by county for 100 counties of North Carolina for 1974-1978. Additional data on number of births by county. Interest in estimating probability of SIDS spatially. Note that this is not an ideal data example for exploring this method and is being used as a temporary exercise to demonstrate proof of concept.

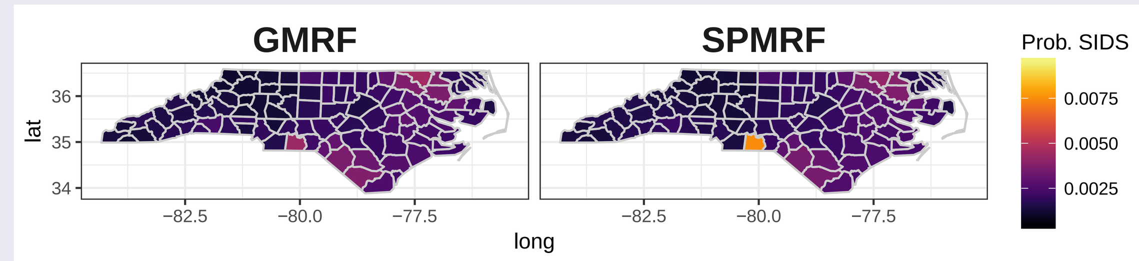
Data



Models

- Let y_i be the number of SIDS deaths and N_i be the number of births for county i over the time period. Then we assume
$$y_i \mid \boldsymbol{\theta} \sim \mathrm{Poisson}(\exp(\theta_i)N_i)$$
- GMRF model
$$\boldsymbol{\theta} \mid \gamma \sim \mathrm{N}(\mu, \mathbf{Q}^{-1})$$
- SPMRF model:
$$\boldsymbol{\theta} \mid \gamma, \boldsymbol{\tau} \sim \mathrm{N}(\mu, \mathbf{Q}(\boldsymbol{\tau})^{-1})$$
$$\tau_{ij} \mid \gamma \sim \mathrm{C}^+(0, \gamma)$$
- For both models $\gamma \sim \mathrm{C}^+(0, 0.1)$

Results



WAIC: 442 (GMRF) vs. 435 (SPMRF)

Anson County: 0.0044 (GMFRF) vs. 0.0075 (SPMRF)

Next steps

- Compare to other locally-adaptive methods with simulations.
- Develop more computationally efficient methods to allow faster inference and larger data sets.
- Investigate how models respond to spatial confounding issues (e.g., Hughes and Haran 2013).
- Extend to continuous space.

References

- Bolin, D. 2014. Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics* 41:557-579.
- Faulkner, JR, and VN Minin. 2017. Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian Analysis.* doi:10.1214/17-BA1050
- Hughes, J, and M Haran. 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society, Series B* 75:139-159.
- Paciorek, CJ, and MJ Schervish. 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17:483-506.
- Yue, Y, and PL Speckman. 2010. Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics* 19:96-116.
- Yue, YR, D Simpson, F Lindgren, and H Rue. 2014. Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis* 9:397-424.
- Wang, Y-X, J Sharpnack, AJ Smola, and RJ Tibshirani. 2016. Trend filtering on graphs. *Journal of Machine Learning Research* 17: 1-41.