

Dimension Reduction and Alleviation of Spatial Confounding for Spatial Generalized Linear Mixed Models

Murali Haran ¹

joint with John Hughes ²

¹Department of Statistics, Penn State University

²Division of Biostatistics, University of Minnesota

Department of Business Statistics and Econometrics,
Guanghua School of Management and Center for Statistical
Science, Peking University, Beijing.

December 2014

What This Talk is About

- ▶ Modeling spatial data on a lattice is challenging.
- ▶ Spatial generalized linear mixed models (SGLMMs) provide a general framework. They are widely used.
- ▶ Shortcomings of SGLMMs: (1) Inference presents difficult computational issues. (2) Parameter interpretation is generally misleading.
- ▶ I will describe an approach that simultaneously resolves both these issues.

Non-Gaussian Spatial Data Example #1

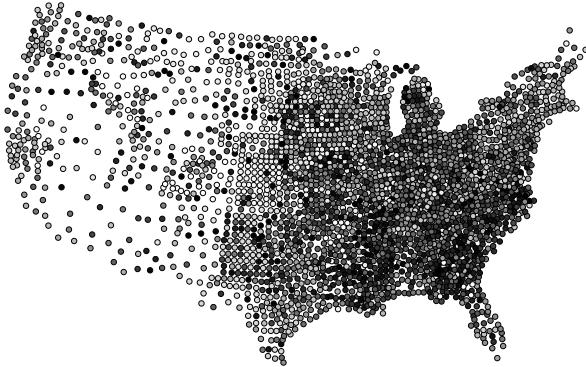
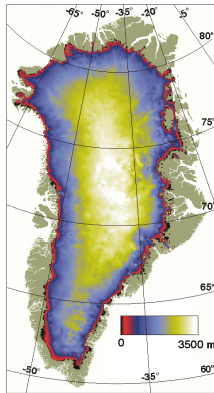


Figure: U.S. infant mortality data by county. $n = 3071$
Ratio of deaths to births, each averaged over 2002-2004.
Darker indicates higher rate.

Non-Gaussian Spatial Data Example #2



Greenland ice sheet thickness (Bamber et al., 2001)

Spatial Data on a Lattice

- ▶ Gaussian and non-Gaussian spatial data are very common and appear in a large number of disciplines.
- ▶ Common lattice data: binary, count, zero-inflated
- ▶ Purpose of the model
 1. regression while adjusting for residual spatial dependence
 2. smoothing the spatial field and “borrowing strength”
- ▶ These models are used widely and have become particularly important in disease epidemiology and ecology.

Spatial Linear Models

- ▶ Spatial process at location \mathbf{s} is $Z(\mathbf{s}) = X(\mathbf{s})\beta + W(\mathbf{s})$.
 - ▶ $X(\mathbf{s})$ are covariates at \mathbf{s} and β is a vector of coefficients.
 - ▶ Model dependence among spatial random variables by imposing it on the errors (the $W(\mathbf{s})$'s).
- ▶ Gaussian Markov Random field (GMRF): Let Θ be the parameters for precision matrix $Q(\Theta)$. Then:

$$\mathbf{z}_{n \times 1} | \Theta, \beta \sim N(\mathbf{X}_{n \times p} \beta_{p \times 1}, Q^{-1}(\Theta))$$

Spatial Linear Models: Dependence

- ▶ $Q = \text{diag}(A\mathbf{1}) - A$ where adjacency matrix A is such that $A_{ij} = 1$ if locations i and j are neighbors, 0 else
- ▶ Implications:
 - ▶ $W(\mathbf{s})$ is conditionally independent of all other W s given its neighbors
 - ▶ uncertainty about $W(\mathbf{s})$ is inversely proportional to its number of neighbors.

Spatial Generalized Linear Mixed Models

Model for Z at location \mathbf{s}_i

1. $Z(\mathbf{s}_i) | \beta, \Theta, W(\mathbf{s}_i), i = 1, \dots, n$, conditionally independent

E.g. $Z(\mathbf{s}_i) | \beta, W(\mathbf{s}_i) \sim \text{Poisson}(\mu(\mathbf{s}_i))$

2. Link function $g(\mu(\mathbf{s}_i)) = X(\mathbf{s}_i)\beta + W(\mathbf{s}_i)$

E.g. $\log(\mu_i) = X(\mathbf{s}_i)\beta + W(\mathbf{s}_i)$

3. Impose dependence: $\mathbf{W} = (W(\mathbf{s}_1), \dots, W(\mathbf{s}_n))^T$

$$p(\mathbf{W} | \tau) \propto \tau^{(n-1)/2} \exp \left(-\frac{\tau}{2} \mathbf{W}' Q \mathbf{W} \right)$$

4. Priors for Θ, β

Inference based on $\pi(\Theta, \beta, \mathbf{W} | \mathbf{Z})$

(Besag et al. (1991), Diggle et al. (1998))

SGLMMs: Challenges

SGLMMs have become very popular even outside mainstream statistics. Flexible models but some drawbacks:

- (1) Confounding between spatial random effects and fixed effects (covariates)
- (2) Computational challenges

Spatial Confounding in SGLMMs

- ▶ $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, orthogonal projection onto $C(\mathbf{X})$
- ▶ $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$, orthogonal projection onto $C(\mathbf{X})$'s orthogonal complement
- ▶ Spectral decomposition to acquire orthogonal bases, $\mathbf{K}_{n \times p}$ and $\mathbf{L}_{n \times (n-p)}$, for $C(\mathbf{X})$ and $C(\mathbf{X})^\perp$. Rewrite:

$$g(\mathbb{E}(Z_i | \beta, W_i)) = \mathbf{X}_i\beta + W_i = \mathbf{X}_i\beta + \mathbf{K}_i\gamma + \mathbf{L}_i\delta.$$

\mathbf{K} is collinear with \mathbf{X} .

This is the source of confounding. Appears to cause variance inflation.

Computing for SGLMMs

MCMC algorithms for SGLMMs are challenging to construct:

- ▶ Spatial random effects: one random effect for each data point. $n + p + 1$ dimensions where n =size of data, p =number of predictors. MCMC is slow per iteration due to high dimensionality
- ▶ Markov chain is slow mixing due to strong cross-correlations among the spatial random effects.

Several attempts to address these issues: Rue and Held (2005), Haran et al. (2003), Haran and Tierney (2010)

Observations

- ▶ Spatial random effects **W** are the cause of confounding issues as well as computational challenges.
- ▶ **W** are just a device to induce dependence. Not intrinsically important.
- ▶ Idea: reparameterize and reduce dimensions of **W**.

Spatial Confounding: Reparameterization Solution

- ▶ Reich, Hodges and Zadnik (2006) propose solution: since \mathbf{K} have no scientific meaning, delete them from the model.
- ▶ $g(\mathbb{E}(Z_i | \beta, \delta)) = \mathbf{X}_i\beta + \mathbf{L}_i\delta$. Prior for random effects δ now

$$p(\delta | \tau) \propto \tau^{(n-p)/2} \exp\left(-\frac{\tau}{2}\delta'\mathbf{Q}^*\delta\right),$$

where $\mathbf{Q}^* = \mathbf{L}'\mathbf{Q}\mathbf{L}$.

- ▶ Corrects issues due to confounding
- ▶ # of parameters reduced (only slightly) from $n + p + 1$ to $n + 1$. Computational challenge remains.
- ▶ RHZ approach does not fully account for underlying graph

Our Sparse Reparameterization

- Represent graph $G = (V, E)$ using \mathbf{A} , $n \times n$ adjacency matrix with entries $\text{diag}(\mathbf{A}) = \mathbf{0}$ and $\mathbf{A}_{ij} = 1\{(i, j) \in E, i \neq j\}$, with $1\{\cdot\}$ an indicator function
- Basic idea inspired by Griffith (2003): augment a generalized linear model with selected eigenvectors of $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{A}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$. This appears in Moran's I statistic (nonparametric measure of spatial dependence),

$$I(\mathbf{A}) \propto \frac{\mathbf{Z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{A}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{Z}}{\mathbf{Z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{Z}},$$

Background for Sparse Reparameterization

- ▶ Griffith's goal: reveal the structure of missing spatial covariates. Our goal: smoothing orthogonal to \mathbf{X}
- ▶ Hence, we replace $\mathbf{I} - \mathbf{1}\mathbf{1}'/n$ with \mathbf{P}^\perp
- ▶ $\mathbf{M}_\mathbf{X}(\mathbf{A}) = \mathbf{P}^\perp \mathbf{A} \mathbf{P}^\perp$, Moran operator for \mathbf{X} with respect to the graph G , appears in numerator of generalized Moran's I :

$$I_\mathbf{X}(\mathbf{A}) \propto \frac{\mathbf{Z}' \mathbf{P}^\perp \mathbf{A} \mathbf{P}^\perp \mathbf{Z}}{\mathbf{Z}' \mathbf{P}^\perp \mathbf{Z}}.$$

Applying the Sparse Reparameterization

- Replacing \mathbf{L} with \mathbf{M} in the RHZ model gives

$$g(\mathbb{E}(Z_i | \beta, \delta)) = \mathbf{X}_i\beta + \mathbf{M}_i\delta.$$

And the prior for the random effects is now

$$p(\delta | \tau) \propto \tau^{q/2} \exp\left(-\frac{\tau}{2} \delta' \mathbf{Q}^{**} \delta\right),$$

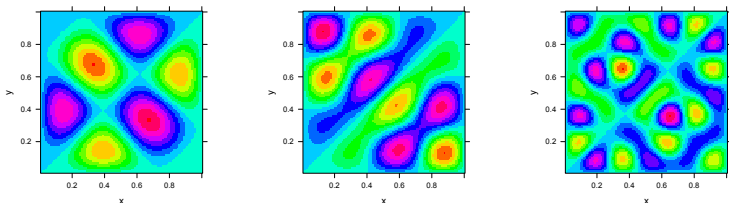
where $\mathbf{Q}^{**} = \mathbf{M}'\mathbf{Q}\mathbf{M}$.

- Corrects issues due to confounding
- Potential for dimension reduction: if we reduce dimensions of \mathbf{M}_i to q , the # parameters is reduced from $n + p + 1$ to $q + p + 1$ (q can be small)

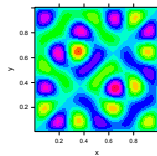
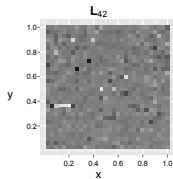
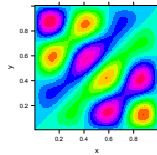
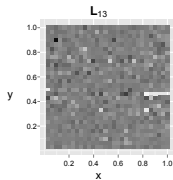
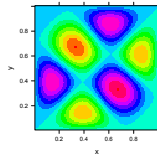
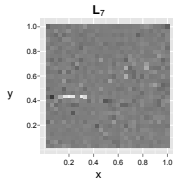
Interpreting the Resulting Reparameterization

- “Tailored” to \mathbf{X} and G : eigenvectors comprise all possible patterns of clustering residual to \mathbf{X} and accounting for G

Some selected basis vectors for the 30×30 lattice.



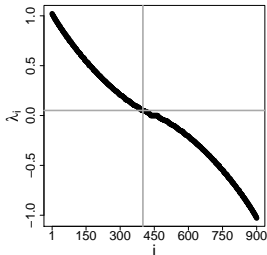
Eigenvectors 7, 13, 42 from RHZ and Moran bases



Interpreting the Resulting Reparameterization

- Positive (negative) eigenvalues correspond to varying degrees of positive (negative) spatial dependence (Boots and Tiefelsdorf, 2000)

The standardized eigenvalues for the 30×30 lattice.



Exploiting the New Parameterization

- ▶ If we assume positive spatial dependence, eigenvectors corresponding to negative spatial dependence (negative eigenvalues) should be removed.
- ▶ Small eigenvalues may not be meaningful. Remove corresponding eigenvectors.
- ▶ Result: much reduced dimensions

Study: Inference for Spatial Binary

30×30 lattice simulated from RHZ model with $\beta_1 = \beta_2 = 1$.

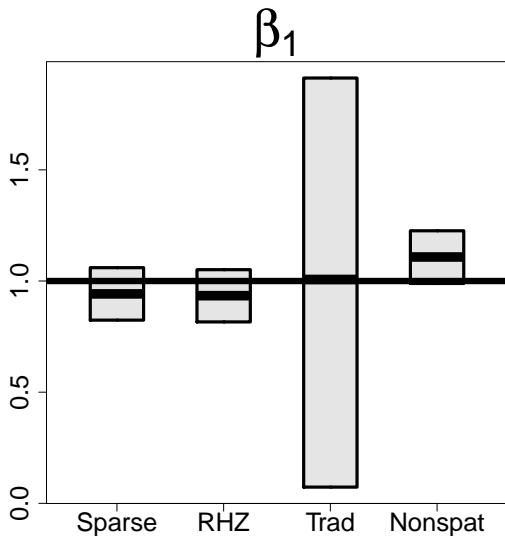
Predictors are the coordinates of unit square.

Model	$\hat{\beta}_1$ CI(β_1)	$\hat{\beta}_2$ CI(β_2)
Sparse	1.080 (0.613, 1.556)	1.130 (0.644, 1.635)
RHZ	1.120 (0.637, 1.606)	1.192 (0.679, 1.713)
Traditional	0.500 (-2.655, 3.616)	-0.605 (-3.698, 2.577)

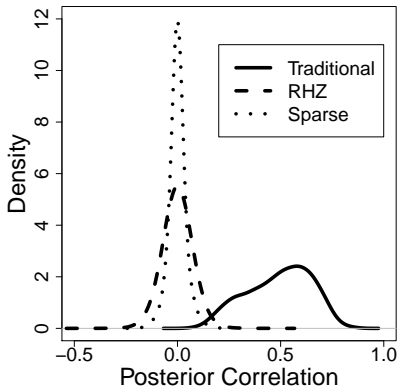
- Point and interval estimates for Traditional are very poor:
95% interval includes 0
- Sparse and RHZ produce similar (good) results

Similar results for Gaussian (linear) and Poisson

Spatial Count Data: Simulation Results



De-correlated Random Effects



Greatly improves efficiency of simple MCMC. No need for elaborate proposals (cf. Held and Rue (2005), Haran et al. (2003), Haran and Tierney (2010)).

Spatial Binary: Computational Efficiency

Model	Dimension	Running Time
Sparse	228	2.5 hours
RHZ	901	18.5 hours
Traditional	903	38.5 hours

- ▶ MCMC algorithm is
 - ▶ faster per iteration (far fewer random effects)
 - ▶ mixes faster (random effects are “decorrelated”)
 - ▶ Far greater speed-ups with much smaller q , e.g. 25-50 is adequate for our examples (we are also being *extremely* careful by running very long chains!)
- Real data example: 14 days (traditional) versus 2-8 hours

Summary

- ▶ SGLMMs provide a very general approach for modeling non-Gaussian spatial data
- ▶ Our sparse approach results in more interpretable regression coefficients
- ▶ We allow for only meaningful spatial dependence and a natural approach to dimension reduction
- ▶ Automated MCMC is computationally efficient, allowing for routine analysis of large data sets

References

- ▶ Besag, York, Mollie (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*
- ▶ Griffith (2003) Spatial Autocorrelation and Spatial Filtering. *Springer*.
- ▶ Reich, Hodges and Zadnik (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*

Hughes, J. and Haran, M. (2013) "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models," *Journal of the Royal Statistical Society (B)*

Software: <http://www.biostat.umn.edu/~johnh/software.html>