A Projection-based Approach for Spatial Generalized Linear Mixed Models

Based on joint work with

Yawen Guan, SAMSI/NC State

John Hughes, U Colorado-Denver

Department of Statistics, North Carolina State University

October 2017

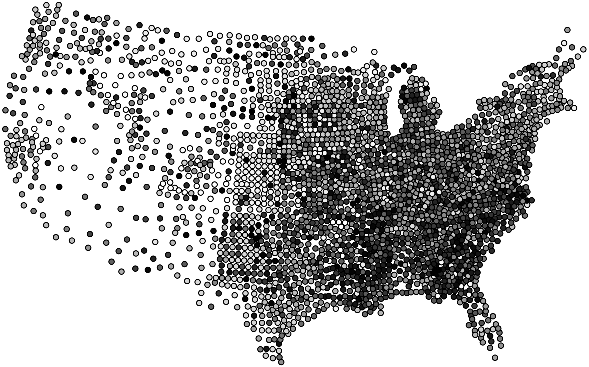
Murali Haran

Department of Statistics, Penn State University

Talk Summary

- Gaussian and non-Gaussian spatial data are common: disease modeling, ecology, climate science, sociology
- Spatial generalized linear mixed models (SGLMMs)
  - Popular for lattice or areal data
    Besag, York, Mollie (1991) $\approx$ 3,000 citations
  - and continuous-domain data
    Diggle et al. (1998) $\approx$ 2,000 citations
- Shortcomings of SGLMMs:
  1. Inference presents difficult computational issues, especially with large data sets
  2. Regression parameter interpretation is unreliable
- I will describe projection-based methods that simultaneously resolve both these issues

US Infant Mortality Data by County

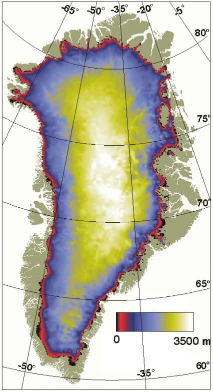


Ratio of deaths to births, each averaged over 2002-2004.

Darker indicates higher rate. $n = 3071$

Question: what factors impact infant mortality?

Greenland Ice Sheet Thickness

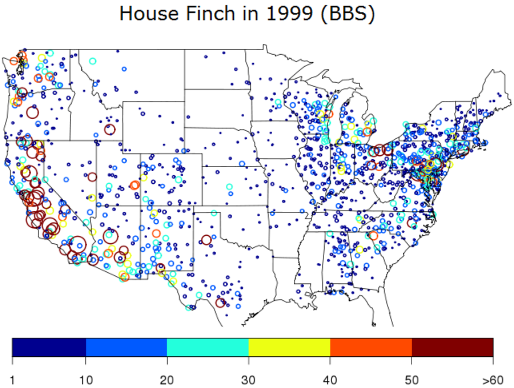


Bamber et al. (2001)

Question: How to interpolate this surface?

How to calibrate (infer parameters for) ice sheet model based on these data?

House Finch Abundances



Pardieck *et al.* 2015. *North American Breeding Bird Survey Dataset 1966 - 2014*

Question: Abundance at unsampled location?

Models for these Data

- Spatial linear mixed models (SLMMs): for Gaussian data
- Spatial generalized linear mixed models (SGLMMs): for non-Gaussian data
- What are these models used for?
  1. interpolation (continuous-domain) or smoothing the spatial field (lattice-domain)
  2. regression while adjusting for residual spatial dependence

Spatial Linear Mixed Models (SLMMs)

- Spatial process at location $\mathbf{s} \in D \subset \mathbb{R}^d$ is

$$Z(\mathbf{s}) = X(\mathbf{s})\boldsymbol{\beta} + W(\mathbf{s})$$

  - $X(\mathbf{s})$ is covariate at $\mathbf{s}$, and $\boldsymbol{\beta}$ is a vector of coefficients
  - Model dependence among spatial random variables by imposing it on $W(\mathbf{s})$, the random effects
- Same framework works for both lattice data and continuous-domain data. Model for $W(\mathbf{s})$
  - Continuous domain: Gaussian process (GP)
  - Lattice data: Gaussian Markov Random field (GMRF)

Gaussian Processes

Infinite dimensional process

$$(W(\mathbf{s}_1), \ldots W(\mathbf{s}_n))^T \mid \Theta \sim N(\mathbf{0}, \Sigma(\Theta))$$

- Covariance often specified via a positive definite covariance function with parameters Θ
- E.g. (stationary) exponential covariance function
- $\Theta = (\sigma^2, \phi, \tau)$

$$\Sigma_{ij}(\Theta) = \text{Cov}(W(\mathbf{s}_i), W(\mathbf{s}_j)) = \sigma^2 \exp(-|\mathbf{s}_i - \mathbf{s}_j|/\phi)$$

Gaussian Markov Random Fields

$$(W(\mathbf{s}_1), \ldots W(\mathbf{s}_n))^T \mid \Theta \sim N(\mathbf{0}, Q(\Theta)^{-1})$$

$Q(\theta)$ is a precision matrix based on a graph that describes a neighborhood structure: adjacencies specify dependence (skip details....)

Inference for Spatial Linear Mixed Models

- MLE involves low-dimensional optimization $\arg\max_{\Theta,\beta} \mathcal{L}(\Theta, \boldsymbol{\beta}; \mathbf{Z})$
- Bayesian inference:
  - Priors for $\Theta, \boldsymbol{\beta}$
  - Inference based on $\pi(\Theta, \boldsymbol{\beta} \mid \mathbf{Z}) \propto \mathcal{L}(\Theta, \boldsymbol{\beta}; \mathbf{Z}) p(\Theta) p(\boldsymbol{\beta})$.
- Markov chain Monte Carlo with low-dimensional posterior

Literature on Computing for Spatial Linear Models

- Likelihood: high-dimensional matrices, $\mathcal{O}(n^3)$ operations.
- Lots of excellent approaches that scale very well
  - Nearest neighbor process (Datta et al., 2016)
  - Predictive process (Banerjee et al., 2008)
  - Multiresolution methods, with parallelizations (Katzfuss, 2017; Katzfuss and Hammerling, 2014)
  - Stochastic PDEs (Lindgren et al., 2011)
  - Lattice kriging (Nychka et al., 2010)

Largely a "solved" problem

Spatial Generalized Linear Mixed Models (SGLMMs)

Model for Z at location $\mathbf{s}_i$

1. $Z(\mathbf{s}_i)|\boldsymbol{\beta}, \Theta, W(\mathbf{s}_i), i = 1, \ldots, n$, conditionally independent
   E.g. $Z(\mathbf{s}_i) \mid \boldsymbol{\beta}, W(\mathbf{s}_i) \sim$Poisson$(\mu(\mathbf{s}_i))$
2. Link function $g(\mu(\mathbf{s}_i)) = X(\mathbf{s}_i)\boldsymbol{\beta} + W(\mathbf{s}_i)$
   E.g. $\log(\mu_i) = X(\mathbf{s}_i)\boldsymbol{\beta} + W(\mathbf{s}_i)$
3. $\mathbf{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n))^T$ modeled as
   - Gaussian Markov random field model (Besag et al., 1991)
   - Gaussian processes (Diggle et al., 1998)
4. Priors for $\Theta, \boldsymbol{\beta}$

Challenges

Challenges posed by spatial generalized linear mixed models (SGLMMs):

(1) Computational challenges

Rue and Held (2002, 2005), Haran (2011)

(2) Confounding between spatial random effects and fixed effects (covariates)

Reich, Hodges, Zadnik (2006), Paciorek (2010)

Problem 1. Computational Challenge

- MLE: low-dimensional optimization of *integrated* likelihood

$$\arg\max_{\Theta,\beta} \int \mathcal{L}(\Theta, \boldsymbol{\beta}, \mathbf{W}; \mathbf{Z}) d\mathbf{W}$$

  High-dimensional integration ($\mathbf{W}$ is high-dimensional)
  MCMC-EM or MCMC-MLE: slow, challenging to implement
  (Zhang, 2002, 2003; Christensen, 2004)

- Bayesian inference based on

$$\pi(\Theta, \boldsymbol{\beta}, \mathbf{W} \mid \mathbf{Z})$$

Computing for SGLMMs

Bayes approach:

- Handle missing data easily
- Combine multiple data sets
- Inference with MCMC is easier (than for MLE)
- But MCMC algorithms is much more challenging
  - MCMC is slow per iteration due to high-dimensional

$$\pi(\Theta, \boldsymbol{\beta}, \mathbf{W} \mid \mathbf{Z})$$

  - Markov chain is slow mixing (need longer Markov chain) due to strong cross-correlations among **W**
  - Can become impractical for large N

MCMC for SGLMMs

- Markov chain is slow mixing (need longer Markov chain) due to strong cross-correlations among **W**
- Block updating schemes may help

$$\boxed{\pi(\mathbf{W} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})} \quad \boxed{\pi(\Theta \mid \boldsymbol{\beta}, \mathbf{W}, \mathbf{Z})} \quad \boxed{\pi(\boldsymbol{\beta} \mid \Theta, \mathbf{W}, \mathbf{Z})}$$

  - Challenging to obtain good proposals for **W**, especially for high-dimensions
  - Computationally expensive per update

Attempts to address these issues: Rue and Held (2005), Christensen et al. (2006), Haran and Tierney (2012)

They do not scale well (*N* typically well under 1000)

Problem 2. Spatial Confounding

- $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, 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:

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

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

Leads to confounding. This leads to variance inflation.

(Reich, Hodges, Zadnik, 2006; Paciorek, 2010)

Sketch of Our Solution

- Culprit: **W** is cause of confounding as well as computational challenges
- **W**: just a device to induce dependence
- Idea: project **W** on random effects δ such that
  - Preserve spatial dependence implied by original **W**
  - δ is low-dimensional
  - δ is less dependent ("cross-correlated")
  - Project orthogonal to space spanned by **X**
- Applies to both Gaussian process and GMRF models
  - GMRF models: projection based on Moran operator which uses neighborhood structure (Hughes and Haran, 2013)
  - GPs and GMRFs: general approach using eigendecomposition (Guan and Haran, 2017)

Sparse Reparameterization for GMRFs

- Delete non-meaningful spatial dependence (weak or negative): “data-based” approach to reduce dimensions
- Regression coefficients are easier to interpret
- Automated MCMC is computationally efficient, allowing for routine analysis of large data sets
- Approach takes advantage of the underlying graph

What should we do in continuous-domain settings (in the absence of a graph)?

SGLMMs with Latent Gaussian Processes

Recall: example model for count data $Z(\mathbf{s}), s \in \mathcal{D} \subset \mathcal{R}^d$.

1. Data model:

$$Z(\mathbf{s}_i) \mid \boldsymbol{\beta}, W(\mathbf{s}_i) \overset{Indep.}{\sim} \text{Poisson}(\mu(\mathbf{s}_i)), i = 1, \ldots, n$$

$$\log(\mu(\mathbf{s}_i)) = X(\mathbf{s}_i)\boldsymbol{\beta} + W(\mathbf{s}_i),$$

2. Process model: impose dependence via Gaussian process

$$\mathbf{W} \mid \sigma^2, \phi \sim N\left(\mathbf{0}, \sigma^2 \Sigma_\phi\right)$$

3. Priors for $\boldsymbol{\beta}, \sigma^2, \phi$

MCMC Inference based on posterior, $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{W} \mid \mathbf{Z})$

Posterior Distribution

$$\pi(\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{W} \mid \mathbf{Z}) \propto$$

$$\prod_i^n f\left(Z(\mathbf{s}_i) \mid \boldsymbol{\beta}, W(\mathbf{s}_i)\right) |\sigma^2 \Sigma_\phi|^{-\frac{1}{2}} \exp\left(-\frac{\mathbf{W}' \Sigma_\phi^{-1} \mathbf{W}}{2\sigma^2}\right) p(\boldsymbol{\beta}, \sigma^2, \phi),$$

where the covariance matrix is specified by the covariance function, for example the i, jth element

$$\Sigma_{ij} = \exp(-|\mathbf{s}_i - \mathbf{s}_j|/\phi)$$

for an exponential covariance function.

Outline of Projection-based Approach

1. Fast approximation to the principal components of Σ_ϕ
   - Approximate first m eigenvectors $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ and eigenvalues $D_m = \text{diag}(\lambda_1, ..., \lambda_m)$
2. Replace n-dimensional $\mathbf{W}$ with $UD_m^{1/2}\boldsymbol{\delta}$

   $\boldsymbol{\delta}$: lower dimensional and $\approx$ independent

   faster and better mixing MCMC algorithm
3. Project $UD_m^{1/2}\boldsymbol{\delta}$ to $C^{\perp}(X)$

   Makes random effects orthogonal to fixed effects

   handles confounding issues
4. Fit the reduced model under Bayesian framework

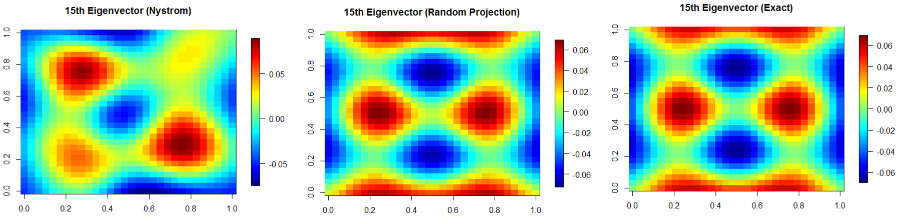
Step 1: Eigendecomposition

For speed we use a fast *approximate* eigendecomposition

Left: deterministic

Center: **random**

Right:exact



- **Random projections** approach used for efficient Gaussian process regression (SLMM) by Banerjee, Tokdar, Dunson (2012)

Step 2: Reducing Dimensions via Projection

- Approximates the leading m eigencomponents of the covariance matrix $K = \Sigma_{\phi}$
- **Replace W with** $UD_m^{1/2}\boldsymbol{\delta}$

Step 3: Orthogonal Projection

- Let $P = X(X^T X)^{-1} X^T$, and $P^{\perp} = I - P$
- Restricted spatial regression: $P\mathbf{W}$ is in span of X. Remove to eliminate confounding [Reich et al., 2006]

$$g\{E(\mathbf{Z} \mid \boldsymbol{\beta}, \mathbf{W}, \sigma^2, \phi)\} = X\boldsymbol{\beta} + \mathbf{W} = X\boldsymbol{\beta} + \cancel{P\mathbf{W}} + P^{\perp}\mathbf{W}$$

- Need adjustment for valid inference [Hanks et al., 2015]

$$\boldsymbol{\beta}^{(k)} = \tilde{\boldsymbol{\beta}}^{(k)} - (X^T X)^{-1} X^T \mathbf{W}^{(k)}$$

- Problem: $P^{\perp}(\mathbf{W}) \sim N(\mathbf{0}, P^{\perp}\Sigma P^{\perp})$ is still high-dim. If X is nxp input matrix, then $P^{\perp}\Sigma P^{\perp}$ has rank n-p.
- Reduce dimension and confounding by $P^{\perp} U D_m^{1/2} \boldsymbol{\delta}$

Step 4: Inference Based on Reparameterizaion

- Spatial generalized linear mixed models
  Usual: inference based on $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{W} \mid \mathbf{Z})$
- Obtain U, D_m of Σ_ϕ
- D_m is m-dim diagonal matrix with $D_{ii} = i^{th}$ eigenvalue
- FRP: replace $\boldsymbol{W}$ with $UD_m^{1/2}\boldsymbol{\delta}$ to approximate SGLMM or
  RRP: replace $\boldsymbol{W}$ with $P^\perp UD_m^{1/2}\boldsymbol{\delta}$ to approximate restricted spatial model
- Reduced Model:

$$g\left\{E(Z_i \mid \boldsymbol{\beta}, U, D_m, \boldsymbol{\delta})\right\} = X_i\boldsymbol{\beta} + (P^\perp UD_m^{1/2})_i\boldsymbol{\delta}$$

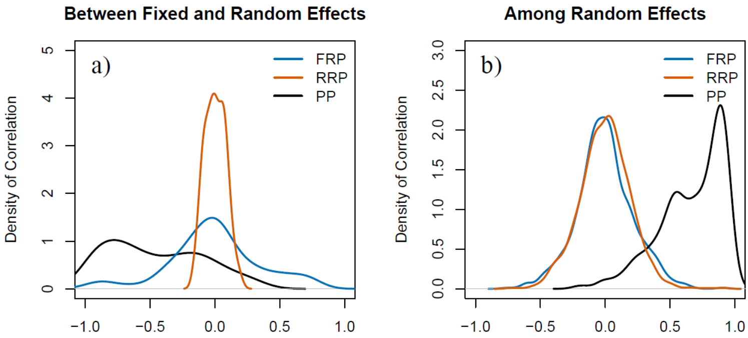
$$\boldsymbol{\delta} \mid \boldsymbol{\theta} \overset{approx}{\sim} N_m(\mathbf{0}, \sigma^2 I)$$

Now: inference based on $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{s}\boldsymbol{\delta} \mid \mathbf{Z})$

Computational Advantages: Improved MCMC Mixing

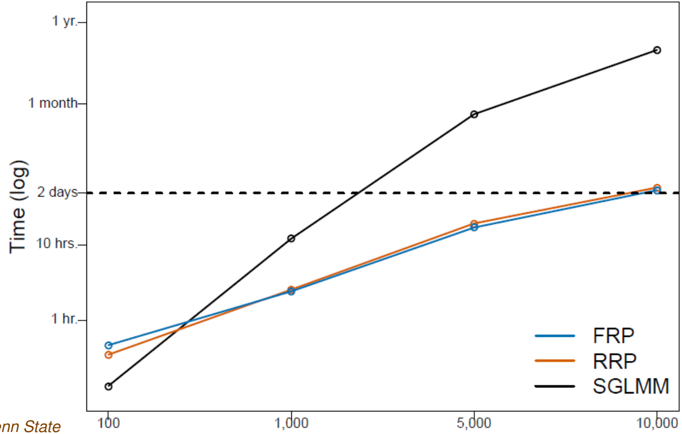
- Alleviate confounding between fixed and random effects
- Reparameterized δ are approximately independent
- De-correlating random effects: better MCMC mixing

Plots of sample cross-correlations



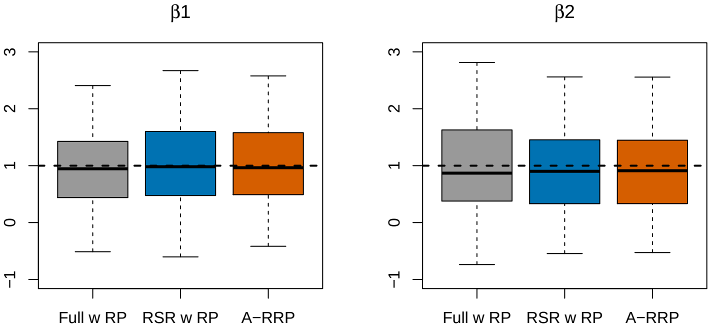
Computational Advantages: Reduced Random Effects

- Can reduce dimension of random effects, e.g. δ to $m << n$ e.g. m = 50, n = 1000.
- Computational complexity: $O(n^2 m)$ versus $O(n^3)$ + mixing improvement (harder to quantify)



Poisson Model Simulation Study: Point Estimation

- Simulate: $\boldsymbol{\beta} = (1, 1)^T$, and Matérn $(\nu, \phi, \sigma^2) = (2.5, 0.2, 1)$



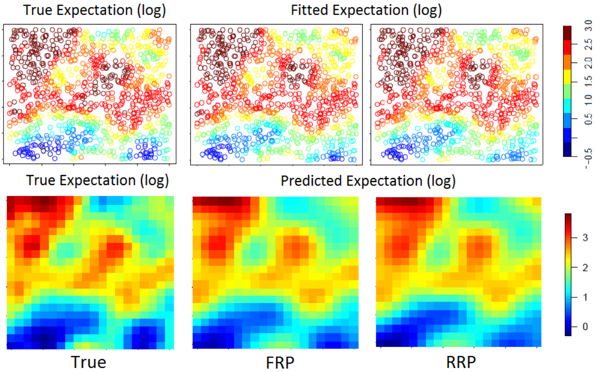
FRP: full model

RRP: restricted model (orthogonalized random effects)

A-RRP: adjusted inference

Poisson Model Prediction Performance

- Simulate n = 1000 spatial count data
- Prediction on 20 x 20 grid using rank = 50



FRP: full model

RRP: restricted model (orthogonalized random effects)

Summary

- Projection-based approach for spatial data
  1. reduces dimensions of posterior distribution
  2. reparameterization improves mixing of MCMC algorithm
  3. adjusts for spatial confounding
  4. simple to implement, mostly "automated"
  5. extends easily to more complex hierarchical settings (not true for multiresolution-type methods even in the spatial linear model case)
- Simulations: good inference and prediction performance
- Caveat: our approach is faster than existing approaches but does not scale to larger data ($n > 10{,}000$ may be problematic)

Acknowledgments

- Yawen Guan, SAMSI/NC State
- John Hughes, U of Colorado-Denver
- SAMSI: discussions with
  - Bo Li (Purdue U.)
  - Doug Nychka (NCAR)
  - Dorit Hammerling (NCAR)
- Support from **NSF-CDSE/DMS-1418090**

Key References

- Guan and Haran (2017), A Computationally Efficient Projection-Based Approach for Spatial Generalized Linear Mixed Models, *arxiv.org*
- Hughes and Haran (2013), Dimension reduction and alleviation of confounding for spatial generalized linear mixed models, *Journal of the Royal Statistical Society (B)*
- Banerjee A, Tokdar, S., Dunson, D. (2011) Efficient Gaussian process regression for large datasets, *Biometrika*
- Reich et al. (2006), Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models *Biometrics*

Frequently Asked Questions (FAQs)

- *Q. Why not use nearest neighbor Gaussian processes? (Datta et al., 2016)*
  - Effective way to reduce matrix calculations via composite likelihood. But does not reduce number of random effects
  - Works well for spatial linear mixed models, not SGLMMs
- *Q. How does your approach compare to the Gaussian predictive process (Banerjee et al., 2008)?*
  - Applicable to SGLMMs, involves dimension-reduction
  - Have to choose “knots” for low-dimensional representation. Non-trivial, far from automated
  - Does not address spatial confounding
  - We address both
  - In simulated examples, we do better with prediction

FAQs

- *Q. Is this necessary when we have the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2008)?*
  - INLA is very fast
  - Does not handle spatial confounding
  - No obvious way to handle complications – additional hierarchy, complicated mean structure (e.g. physical model); accuracy of approximation may also be suspect
- *Q. Relationship to fixed rank approaches?*
  - If we fixed covariance parameters, this is a fixed rank approach with fixed eigenvectors/eigenfunctions as basis
  - Eliminating small scale variations can impact SLMMs (Stein, 2014), but less impact in SGLMMs