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, Duke University

October 2017

Murali Haran

Department of Statistics, Penn State University

- Benjamin Disraeli

  “Talk to a man about himself and he’ll listen to you for hours.”

- Benjamin Disraeli

  "Talk to a man about himself and he'll listen to you for hours."

- My academic version

  "Cite the work of academics, and they might listen to you for 50 minutes."

- Benjamin Disraeli

  "Talk to a man about himself and he'll listen to you for hours."
- My academic version

  "Cite the work of academics, and they might listen to you for 50 minutes."
  - Banerjee, A. **Tokdar**, **Dunson** (2013) Efficient Gaussian process regression for large datasets

- Benjamin Disraeli
  "Talk to a man about himself and he'll listen to you for hours."
- My academic version
  "Cite the work of academics, and they might listen to you for 50 minutes."
  - Banerjee, A. **Tokdar**, **Dunson** (2013) Efficient Gaussian process regression for large datasets
  - Banerjee, **Gelfand**, Finley, Sang (2008) Gaussian predictive process models for large spatial data sets
  - Datta, A., Banerjee, S., Finley, A. O., **Gelfand** (2015) Hierarchical nearest neighbor Gaussian process models for large geostatistical datasets

- Benjamin Disraeli

  "Talk to a man about himself and he'll listen to you for hours."

- My academic version

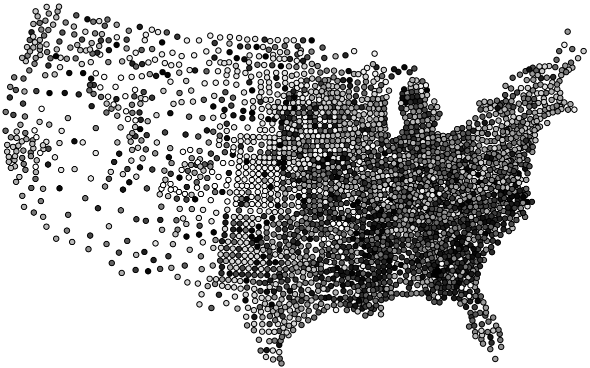
  "Cite the work of academics, and they might listen to you for 50 minutes."

  - Banerjee, A. **Tokdar**, **Dunson** (2013) Efficient Gaussian process regression for large datasets
  - Banerjee, **Gelfand**, Finley, Sang (2008) Gaussian predictive process models for large spatial data sets
  - Datta, A., Banerjee, S., Finley, A. O., **Gelfand** (2015) Hierarchical nearest neighbor Gaussian process models for large geostatistical datasets
  - **Berger**, Liseo, **Wolpert** (1999) Integrated likelihood methods

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

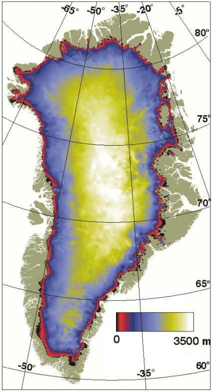


Ratio of deaths to births, each averaged over 2002-2004.

Darker indicates higher rate. $n = 3071$

Question: what factors impact infant mortality?

Greenland Ice Sheet Thickness

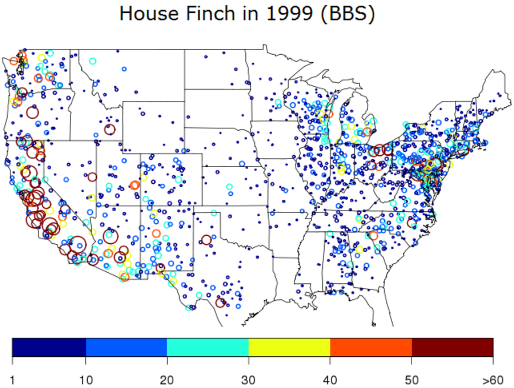


Bamber et al. (2001)

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 $\{W(\mathbf{s}) : \mathbf{s} \in D\}$ such that

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

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

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

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
  - Multiresolution methods, with parallelizations (Katzfuss, 2017; Katzfuss and Hammerling, 2014)
  - Nearest neighbor process (Datta et al., 2016)
  - Random projections (Banerjee, A., Tokdar, Dunson, 2013)
  - Stochastic PDEs (Lindgren et al., 2011)
  - Lattice kriging (Nychka et al., 2010)
  - Predictive process (Banerjee, Gelfand, Finley, Sang 2008)

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 $\Theta, \boldsymbol{\beta}$

Commonly embedded within hierarchical models (cf. Banerjee, Carlin, Gelfand, 2014)

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

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

  (Nice study of integrated likelihood methods: Berger, Liseo, Wolpert, 1999)

  High-dimensional integration (**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 are not easy/scalable
  - 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 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. E.g. blocks:

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

  - 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 (problem for $N > 1000$)

Problem 2. Spatial Confounding

- Let $P = X(X^T X)^{-1} X^T$, and $P^{\perp} = I - P$

$$g\{E(\mathbf{Z} \mid \beta, \mathbf{W}, \Theta\} = X\beta + \mathbf{W} = X\beta + \boxed{P\mathbf{W}} + P^{\perp}\mathbf{W}$$

- $P\mathbf{W}$ is in span of X
- Basic regression issue: multicollinearity

Leads to variance inflation, unstable estimates of β

(Hodges and Reich 2010; Paciorek, 2010)

Hints of the symptom, without diagnosis, by others (e.g. Diggle, 1994)

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

- Regular approach implies unintended/undesirable dependence structure (cf. Wall, 2004)
- Our approach
  - Deletes non-meaningful spatial dependence (weak or negative): “data-based” approach to reduce dimensions
  - Faster inference *and* a better model
- 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:

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

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

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 **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

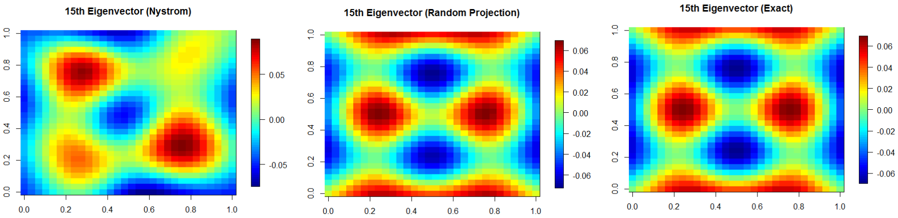
Step 1: Eigendecomposition

For speed we use a fast *approximate* eigendecomposition

Left: deterministic approximation

Center: **random approximation**

Right: exact eigendecomposition



- **Random projections** used in Banerjee, Tokdar, Dunson (2013); also Sarlos (2006), Halko et al. (2009)

Step 2: Reducing Dimensions via Projection

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

Step 3: Projection to Handle Confounding

- Let $P = X(X^T X)^{-1} X^T$, and $P^\perp = I - P$
- Recall: $P\mathbf{W}$ is in span of X, causes confounding
- Solution: Remove it

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

[Reich et al., 2006]

- High-dimensional $P^\perp \mathbf{W} \sim N(\mathbf{0}, P^\perp \Sigma P^\perp)$
  If X is nxp input matrix, then $P^\perp \Sigma P^\perp$ has rank n-p
- Only reduces dimensions from n to $n - p$
- Instead: 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 **W** with $UD_m^{1/2}\boldsymbol{\delta}$ to approximate SGLMM or
  RRP: replace **W** with $P^\perp UD_m^{1/2}\boldsymbol{\delta}$ to approximate restricted spatial model
- Reduced Model:

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

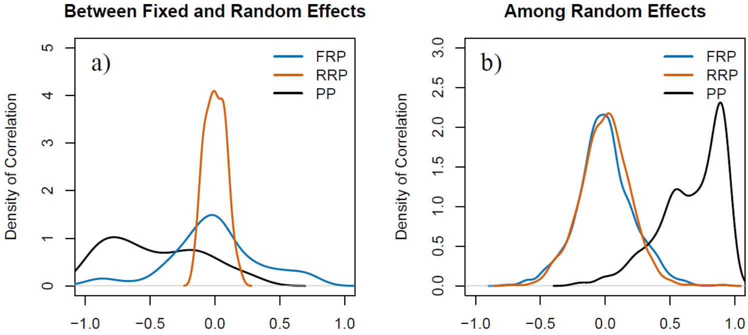
$$\boldsymbol{\delta} \mid \ldots \overset{approx}{\sim} N_m(\mathbf{0}, \sigma^2 I)$$

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

Computational Advantages: Improved MCMC Mixing

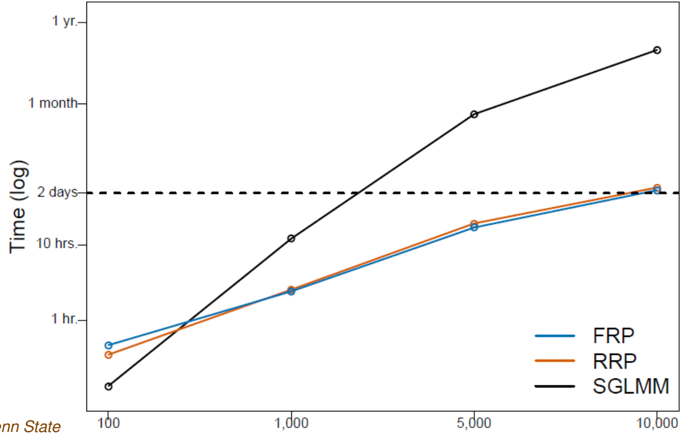
- Alleviate confounding between fixed and random effects
- Reparameterized δ are approximately independent
- De-correlating random effects: better MCMC mixing

Plots of sample cross-correlations



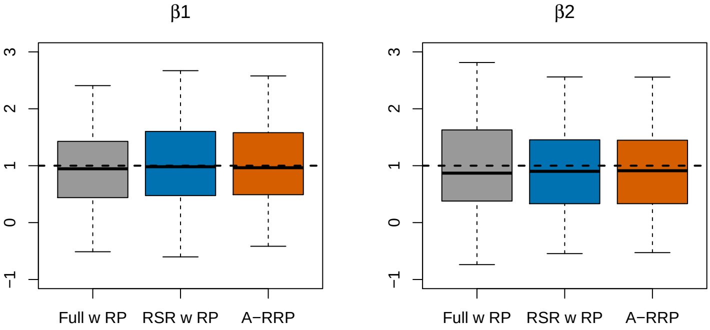
Computational Advantages: Reduced Random Effects

- Can reduce dimension of random effects, $\boldsymbol{\delta}$ 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: $\beta = (1, 1)^T$, and Matérn $(\nu, \phi, \sigma^2) = (2.5, 0.2, 1)$



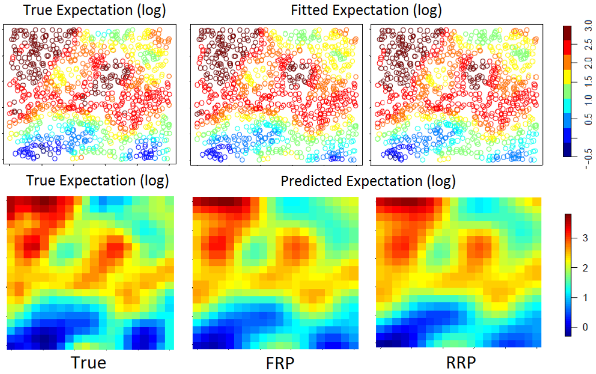
FRP: full model

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 + better MCMC mixing
  2. adjusts for spatial confounding
  3. simple to implement, mostly “automated”
  4. good inference and prediction performance
  5. other approaches (nn-GP, random-proj, Multi-Re) are better than ours for basic linear model; we are better for SGLMMs
  6. extends easily to more complex hierarchical settings (not true for multiresolution-type methods even in the spatial linear model case)
- Caveat: Have not studied method for $n > 10,000$
  - For fixed m, computational cost grows with n (mostly) due to eigendecomposition. Address via (i) discretization of space/pre-computing and (ii) new algorithms

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... *Journal of the Royal Statistical Society (B)*
- Banerjee A, Tokdar, S., Dunson, D. (2013) 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*
- Haran (2011) Gaussian random field models for spatial data, *Handbook of MCMC*

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
  - Random effects are of dimension N so not clear how to extend to 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
  - 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

APPENDIX

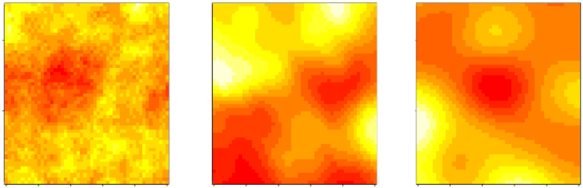
Gaussian Process for Dependence and Interpolation

- A Gaussian process is an infinite-dimensional random process, any finite-dimension of which is a multivariate normal.
- Matérn covariance function describes dependence, e.g.

$$\nu = 0.5, \quad C(h) = \sigma^2 \exp(-\tfrac{|h|}{\phi}) \text{ (Exponential)}$$

$$\nu = 2.5, \quad C(h) = \sigma^2 \left(1 + \frac{\sqrt{5}|h|}{\phi} + \frac{5|h|^2}{3\phi^2}\right) \exp(-\frac{\sqrt{5}|h|}{\phi})$$

$$\nu = \infty, \quad C(h) = \sigma^2 \exp(-\tfrac{|h|^2}{2\phi^2}) \text{ (Square exponential)}$$



Our Sparse Reparameterization

- Represent graph $G = (V, E)$ using $\mathbf{A}$, $n \times n$ adjacency matrix with entries $\mathrm{diag}(\mathbf{A}) = \mathbf{0}$ 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 **L** with **M** in the RHZ model gives

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

  And the prior for the random effects is now

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

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

- Corrects issues due to confounding
- Dimension reduction: if $\mathbf{M}_i$ reduced to q dimensions
  # parameters $q + p + 1 << n + p + 1$ if q is small

Study: Inference for Spatial Binary

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

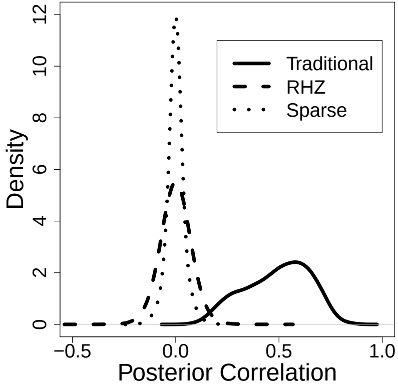
Predictors are the coordinates of unit square.

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

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

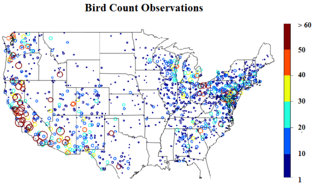
Model	Dimension	Running Time
Sparse	228	2.5 hours
RHZ	901	18.5 hours
Traditional	903	38.5 hours

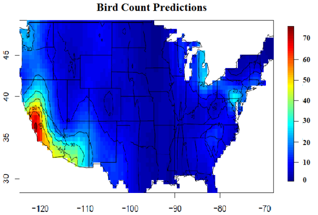
- MCMC algorithm is
  - faster per iteration (far fewer random effects)
  - mixes faster (random effects are "decorrelated")
- Far greater speed-ups with much smaller q, e.g. 25-50 is adequate for our examples (we are also being *extremely* careful by running very long chains!)

  Real data example: 14 days (traditional) versus 2-8 hours

Interpolated Bird Counts

- Approximate the SGLMM with only the intercept term.
- Computation time is about 7 hours,
- Small bird counts in the center and most of the East Coast
- Large counts centered near New York area and the West





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

Outline of Projection-based Approach

1. Fast (approximate) eigendecomposition of Σ_ϕ:
   1.1 Low-distortion embedding of Σ_ϕ,
   1.2 Approximate first m eigenvectors $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ and eigenvalues $D_m = \text{diag}(\lambda_1, ..., \lambda_m)$ via Nyström method. [Banerjee et al., 2012] used a similar algorithm to approximate Σ_ϕ in Gaussian process regression
2. Replace n-dimensional **W** with $UD_m^{1/2}\boldsymbol{\delta}$
   $\boldsymbol{\delta}$: lower dimensional, components $\approx$independent
3. Project $UD_m^{1/2}\boldsymbol{\delta}$ to $C^\perp(X)$
   - Makes random effects orthogonal to fixed effects
4. Fit the reduced model under Bayesian framework.

Gaussian Markov Random Fields

$$W(\mathbf{s}_i) \mid W(\mathbf{s}_{-i}) \sim N\left(\frac{\sum_{j:j\sim i} W(\mathbf{s}_j)}{n_i}, \frac{1}{n_i\tau}\right)$$

where n_i is number of neighbors of ith region and $j \sim i$ means i, j are neighboring regions

- This specifies $Q(\tau)$, a precision matrix

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

$Q = \text{diag}(A\mathbf{1}) - A$, where adjacency matrix A is such that $A_{ij} = 1$ if locations i and j are neighbors, 0 else

Spatial Confounding: Reparameterization Solution

- Since $\mathbf{K}$ is collinear, delete it from model
- $g(\mathbb{E}(Z_i \mid \boldsymbol{\beta}, \boldsymbol{\delta})) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{L}_i \boldsymbol{\delta}$. Random effects distribution $\boldsymbol{\delta}$

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

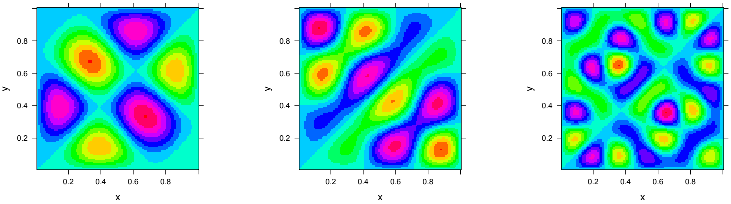
  where $\mathbf{Q}^* = \mathbf{L}'\mathbf{Q}\mathbf{L}$.
- Corrects issues due to confounding
- # of parameters reduced (only slightly) from $n + p + 1$ to $n + 1$. Computational challenge remains.

Reich, Hodges, Zadnik (2006)

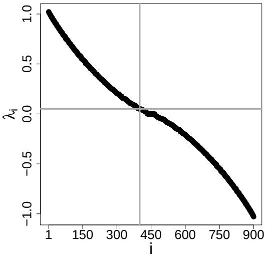
Sketch for Gaussian Markov Random Fields

- "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.



Interpretation: Standardized eigenvalues



- Positive (negative) eigenvalues correspond to degrees of positive (negative) dependence (Boots and Tiefelsdorf, 2000)
- Idea: Remove eigenvectors corresponding to negative (unwanted dependence) or small eigenvalues (noise)

Spatial Count Data: Simulation Results