Projection-based Methods for Hierarchical Spatial Models

Collaborators:

Ben Seiyon Lee, Penn State University

Yawen Guan, University of Nebraska

Big Data Institute, Oxford University.

September 2019

Murali Haran

Department of Statistics, Penn State University

Talk Summary

- Hierarchical spatial models are applicable to many disciplines, including disease modeling, ecology, climate science, sociology
  - Popular for lattice or areal data
    Besag, York, Mollie (1991) $\approx$ 3,000 citations
  - Continuous-domain or point-level (geostatistical) data
    Diggle et al. (1998) $\approx$ 3,000 citations
- Challenges:
  1. Computational
  2. Regression parameter interpretation
- This talk is on projection-based methods to help address these issues

The “Punchline”

- Hierarchical spatial models use high-dimensional latent (unobservable) variables to describe dependence
- Computational challenges can get in the way of analyses
- It is possible to work with a much lower dimensional representation of the latent variables, which greatly speeds up computing

Key References

- Banerjee, Carlin, and Gelfand (2014): Hierarchical Modeling and Analysis for Spatial Data (2014), Chapman & Hall/CRC Press
- Haran (2011) Gaussian random fields for spatial data
- Hughes and Haran (2013) Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models
- Guan and Haran (2018): A computationally efficient projection-based approach for spatial generalized linear mixed models, *J of Computational and Graphical Statistics*
- Lee and Haran (2019): A Discretized Projection-based Approach for Hierarchical Spatial Models (*in prep*)

Outline

Hierarchical Spatial Models

The Computing Challenge

Projection-based Approach to Hierarchical Spatial Modeling

Extension to Continuous Domain Models

Examples

Outline

Hierarchical Spatial Models

The Computing Challenge

Projection-based Approach to Hierarchical Spatial Modeling

Extension to Continuous Domain Models

Examples

Hierarchical Framework

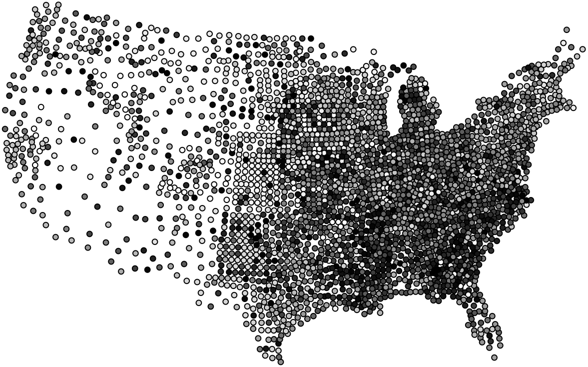
(cf. Mark Berliner, 1994)

1. **Prior** $p(\theta)$: describes assumptions/uncertanties about parameters (θ) of the model.
2. **Process** model ($g(X|\theta)$): describes the model for the process of interest, e.g. a dynamical system that you cannot observe directly
3. **Data** model ($f(Y|X,\theta)$): probability model that describes the observation process, e.g. measurement error and other complications.

Systematic approach to scientific modeling:

(1) $Y|X$, (2) $X|\theta$, (3) θ

Areal Data Example



US infant mortality data by county (Yang et al., 2008)

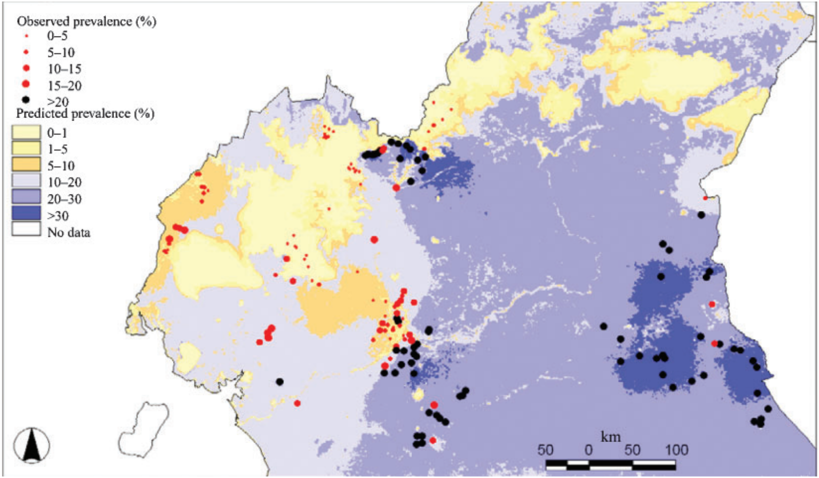
Ratio of deaths to births, each averaged over 2002-2004.

Darker indicates higher rate. $n = 3{,}071$

Question: which factors impact infant mortality?

Point-level Data Example

Loa Loa Prevalence (Diggle et al., 2007)



Hierarchical Spatial Models

- Process model typically used to describe spatial dependence
- Can build on this structure to model very complicated phenomena, account for missing data, integrate multiple data sets
- Common: describe spatial dependence for non-Gaussian data

(cf. Banerjee, Carlin, Gelfand, 2014)

Spatial Linear Mixed Models (SLMMs)

Consider some spatial domain $D \subset \mathbb{R}^2$

- Spatial data at location $\mathbf{s} \in D$ 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.
  - **Process model**: Model dependence among spatial random variables by imposing it on $W(\mathbf{s})$s, the random effects.
- Model for spatial dependence for $\{W(\mathbf{s}), \mathbf{s} \in D\}$
  - Areal data: Gaussian Markov Random field (GMRF)
  - Point-level (geostatistics): Gaussian process (GP)

Spatial Linear Mixed Models for Areal Data

Data on a lattice/aggregate level

- Gaussian Markov random field

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

  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 Linear MMs for Point-level Data

- Model dependence via a Gaussian process:

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

where $\Sigma_{ij} = Cov(W(\mathbf{s}_i), W(\mathbf{s}_j)) = C(||\mathbf{s}_i - \mathbf{s}_j||)$, is specified via a positive definite covariance function with covariance function parameters Θ.

E.g. exponential covariance function with parameters $\Theta = (\sigma^2, \phi, \tau)$.

Spatial Linear Mixed Models: Inference

For both lattice and continuous-domain data:

- Maximum likelihood: maximize $\mathcal{L}(\Theta, \boldsymbol{\beta}; \mathbf{Z})$ w.r.t. $\Theta, \boldsymbol{\beta}$
- Optimization problem is low-dimensional
- Bayesian inference:
  - Priors for $\Theta, \boldsymbol{\beta}$
  - Inference based on $\pi(\Theta, \boldsymbol{\beta} \mid \mathbf{Z}) \propto \mathcal{L}(\Theta, \boldsymbol{\beta}; \mathbf{Z}) p(\Theta) p(\boldsymbol{\beta})$.
- Low-dimensional posterior: use Markov chain Monte Carlo
- Computing: likelihood evaluations involve high-dimensional matrices, n^3 operations
  - GMRFs: sparse matrices $\Rightarrow$ computationally efficient
  - GPs: lots of research, e.g. reduced-rank methods

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

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

4. Priors for $\Theta, \boldsymbol{\beta}$

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

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

Outline

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:

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

  K is collinear with **X**.

Leads to confounding. This appears to cause variance inflation $\Rightarrow$ harder to trust inference about $\boldsymbol{\beta}$

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.

$\Rightarrow$ difficult to construct good algorithm + takes too long to run

Rich Literature on Fast Computing

- Many ideas, most designed for linear spatial models
  - Multiresolution methods, with parallelizations (Katzfuss, 2017; Katzfuss and Hammerling, 2014)
  - Nearest neighbor process (Datta et al., 2016)
  - Random projections (Banerjee, A., Tokdar, Dunson, 2013)
  - Lattice kriging (Nychka et al., 2010)
- A few approaches that work well for SGLMMs:
  - Predictive process (Banerjee, Gelfand, Finley, Sang 2008)
    - Works well, very general. *Finding knots etc. is challenging.*
  - Stochastic PDEs + INLA (Lindgren et al., 2011)
    - Fast! Approximately integrates out $\mathbf{W}$, numerical integration
    - Models missing: ordinal, multivariate spatial; data model involving numerical model, more flexible models...

Outline

Hierarchical Spatial Models

The Computing Challenge

Projection-based Approach to Hierarchical Spatial Modeling

Extension to Continuous Domain Models

Examples

Sketch of Our Solution

Observation:

- Spatial random effects **W** are the cause of confounding issues as well as computational challenges.
- **W**: just a device to induce dependence.

Suggests a solution:

- Idea: project **W** to lower dimensional random effects δ
  - Preserve spatial dependence implied by original **W**
  - 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
  - GPs and GMRFs: general approach using random projections

Spatial Confounding: Reparameterization Solution

- Since **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)

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

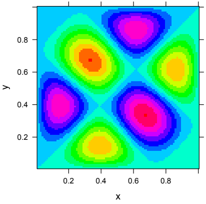
- Corrects issues due to confounding
- Dimension reduction: if $\mathbf{M}_i$ reduced to q dimensions

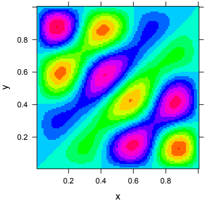
  # parameters $q + p + 1 << n + p + 1$ if q is small

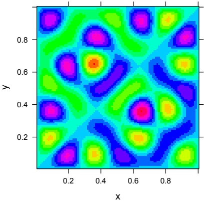
Interpreting the Resulting Reparameterization

- "Tailored" to **X** and G: eigenvectors comprise all possible patterns of clustering residual to **X** and accounting for G

Some selected basis vectors for the 30×30 lattice.



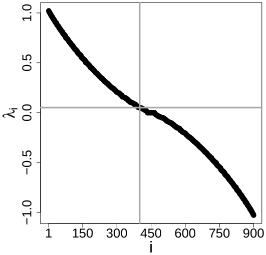




Interpreting the Resulting Reparameterization

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

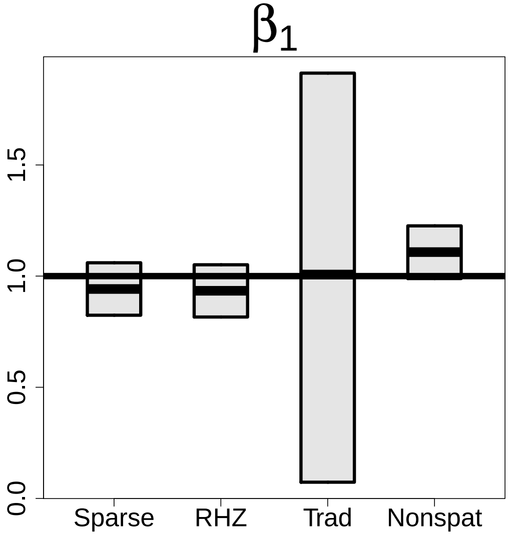
The standardized eigenvalues for the 30×30 lattice.



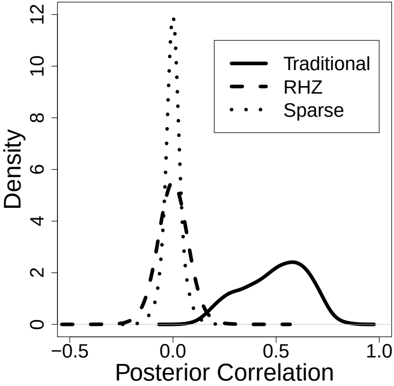
Exploiting the New Parameterization

- 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

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 $\approx$ 2 hours

Code for Projection-based Approach

`R package` ngspatial available on `CRAN`

Outline

Hierarchical Spatial Models

The Computing Challenge

Projection-based Approach to Hierarchical Spatial Modeling

Extension to Continuous Domain Models

Examples

Outline of Projection for Continuous Domain

- Methodology discussed so far: applies to hierarchical spatial models for areal data
- What to do when data are continuous-domain ("geostatistics")?
- Guan and Haran (2018) suggests analogous approach: projection of $\mathbf{W}$ via "random projections" algorithm
  - Fast, but not as fast as previous approach
  - No theoretical justification for approach
- Motivates new PICAR approach (Lee and Haran, 2019)

Sketch of PICAR

PICAR = Projected Intrinsic Conditional Autoregression

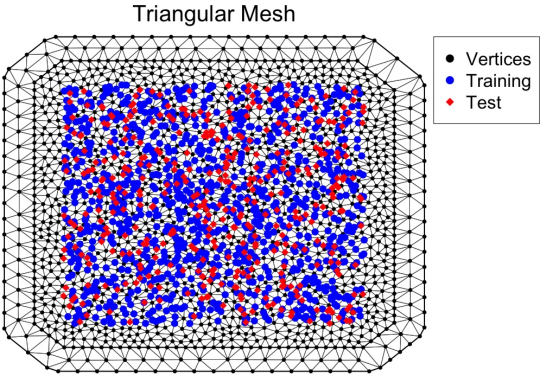
1. **Mesh Construction:** Subdivide spatial domain into a set of non-intersecting triangles (Lindgren et al., 2011)
2. **ICAR model:** Model the mesh vertices as an intrinsic Gaussian Markov random field
3. **Projection:** Reduce dimension + de-correlate spatial random effects (Hughes and Haran, 2013; Griffith, 2003)
4. **Interpolation:** Interpolate the latent continuous Gaussian random field using mesh vertices + basis functions (Lindgren et al., 2011)

Steps 1 and 2: convert point-level model to areal (lattice) model. Rest same as before

Mesh Construction

Goal: Divide spatial domain into a set of non-intersecting irregular triangles (Delaney Triangulation) using `R-INLA`

- Random Effects: $\mathbf{W} \in \mathbb{R}^n$
- Vertices of mesh: $\tilde{\mathbf{W}} \in \mathbb{R}^m$, where $m > n$



Intrinsic Conditional Autoregressive (ICAR) Model

Objective: Model the latent intrinsic Gaussian Markov random field using mesh vertices $\tilde{W}(s)$.

Intrinsic Gaussian Markov Random Field:

$$\mathbf{W}|\tau \sim N(0, [\tau(\mathbf{D} - \mathbf{W})]^{-1}), \tag{1}$$

where:

- τ is the precision parameter
- $\mathbf{W} \in \mathbb{R}^{m \times m}$ is the neighborhood matrix where $\mathbf{W}_{ij} = 1$ when vertices i and j share an edge and $\mathbf{W}_{ij} = 0$ otherwise.
- $\mathbf{D} \in \mathbb{R}^{m \times m}$ where $\mathbf{D}_{i,i}$ = the number of neighbors for vertex i and 0 on the off-diagonals.

Dimension-Reduction + De-correlation

- Objective: Reduce dimensions of random effects (**W**)
- Moran's I

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

  where **Z** are spatial random variables, **P** is precision matrix
- Moran's Operator $(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{n})\mathbf{P}(\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{n})$
- Dimension-Reduction
  1. Generate Moran's basis $\mathbf{M} \in \mathbb{R}^{m \times q}$ containing the first $q << m$ eigenvectors of the Moran's operator
  2. Approximate the latent GMRF as $\tilde{\mathbf{W}} \approx \mathbf{M}\delta$, where new random effects are δ are q-dimensional and decorrelated

Moran's Basis: Eigenvectors of Moran's Operator

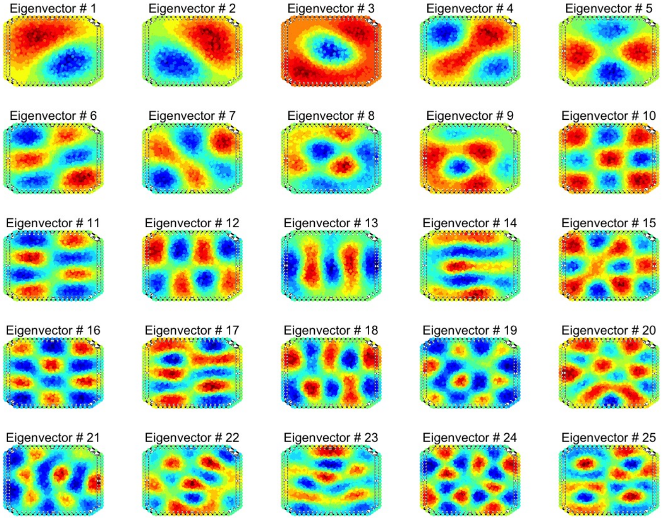
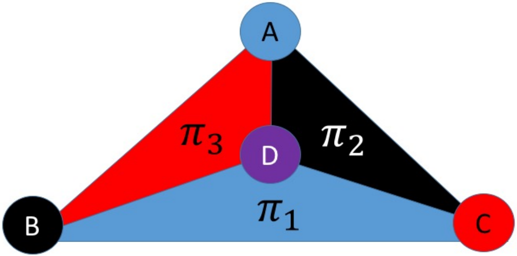


Figure: Eigencomponents

Interpolation Within the Mesh

Basis Functions for Interpolation

- $\mathbf{A}$ is an $n \times m$ projector matrix containing the basis coefficients
- Each row $\mathbf{A}$ corresponds to the location of an observation and each column corresponds to a mesh vertex.
- **Interpolation:** $\mathbf{W} \approx \mathbf{A}\tilde{\mathbf{W}}$, where $\mathbf{W} \in \mathbb{R}^n$ are the spatial random effects and $\tilde{\mathbf{W}} \in \mathbb{R}^m$ are the mesh vertices.



$$D \approx (\pi_1 \text{ x } A) + (\pi_2 \text{ x } B) + (\pi_3 \text{ x } C)$$

Hierarchical Model

Data Model:

$$Z(s) \sim \prod_{i=1}^{n} p(\eta(s)|\beta,\delta)$$

$$\eta(s) = g(E[Z(s)|\beta,\delta]) = X(s)\beta + [AM\delta](s),$$

A is projector matrix, M is Moran's basis, δ are low-dimensional random effects.

Process Model:

$$\delta \sim \mathcal{N}(0, (\tau M(D-W)M)^{-1}),$$

where $(D-W)$ is the precision matrix of the ICAR model.

Priors: $\tau \sim IG(\alpha_\tau, \beta_\tau),\ \beta \sim N(0, \Sigma_\beta)$

Outline

Hierarchical Spatial Models

The Computing Challenge

Projection-based Approach to Hierarchical Spatial Modeling

Extension to Continuous Domain Models

Examples

Simulation Study

Overview:

- Generate 100 spatial count samples with locations on the unit domain $[0, 1]^2$
- $n_{mod} = 1000$ observations to fit model + $n_{cv} = 400$ for cross-validation.
- **Fixed Effects:** $\beta = (1, 1)^T$
- **Random effects $W(s)$:** Generated using the Matérn covariance function with parameters $\nu = 2.5$, $\sigma^2 = 1$, and $\phi = 0.2$.

Model Fitting:

- **Priors:** $\beta \sim N(\mathbf{0}, 100I)$ and $\tau \sim \text{Gamma}(0.5, 2000)$.
- **Vary Moran's Basis Dimensions:** $m = 10, 30, 50, 100, 200, 300$.
- Comparison with INLA

Simulated Example Results

Table: Timing based on $\approx$ 250k iterations of the MCMC algorithm.

Dim	β_1	β_2	CVMPSE	Time (Minutes)
10	0.87 (0.76,0.98)	1.05 (0.94,1.16)	1.34	7.91
50	0.98 (0.86,1.1)	1.05 (0.95,1.17)	1.06	8.57
100	0.99 (0.86,1.11)	1.07 (0.96,1.19)	0.91	9.64

Simulation Study Results: Point Estimates

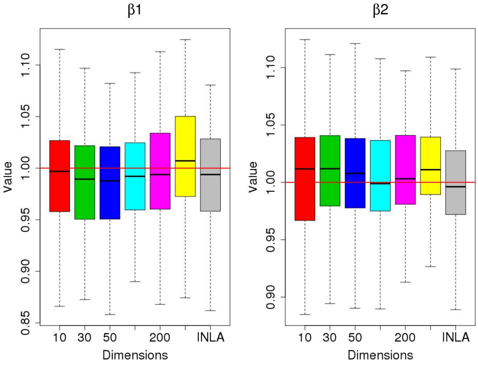


Figure: Boxplots Poisson

Figure: Boxplots illustrating inference for the fixed effects β_1 and β_2 for the 100 simulated Poisson data sets.

Simulation Study Results: Coverage

Table: Coverage Probabilities for 95% credible intervals

Dim	β_1	β_2
10	0.86	0.87
50	0.92	0.94
100	0.95	0.96

The Last Slide

- Project $\boxed{Y \mid Z,\ Z \mid \mathbf{W},\ \mathbf{W} \mid \theta}$ down to: $\boxed{Y \mid Z,\ Z \mid \boldsymbol{\delta},\ \boldsymbol{\delta} \mid \theta}$, where $\boldsymbol{\delta}$ is low-dimensional, less correlated
- Latent Gaussian Markov random field models
  - Moran projections to get $\boldsymbol{\delta}$
- Latent Gaussian process models
  1. Mesh-based discretization
  2. Moran projections to get $\boldsymbol{\delta}$
- Fast, automated, good approximation
- Reduces dimensions + improves mixing of MCMC

Acknowledgments

- Students (current and former): Ben Lee, Yawen Guan, John Hughes, Jaewoo Park
- Papers on `arxiv` online
- Support from National Institutes of Health under Award Number R01GM123007. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.