Dimension Reduction and Spatial Confounding in Models for Non-Gaussian Spatial Data

Murali Haran

(joint work with John Hughes)

Department of Statistics

Penn State University

Workshop on Environmetrics, National Center for Atmospheric Research

October 2010

Non-Gaussian Spatial Models

- Non-Gaussian spatial data are very common and appear in a large number of disciplines, e.g. ecology, epidemiology, forestry.
- Examples: binary, count, zero-inflated.
- Two classes of models:
  - **Automodels** date back to Ising (1941) in statistical mechanics. In statistics, general formulation: Besag (1972, 1974, 1986), Kaiser and Cressie (2000), many others.
    Special case: Gaussian Markov random fields (GMRFs).
  - **Spatial generalized linear mixed models (SGLMMs)**: Besag, York, Mollie (1991), Diggle et al. (1998).
    Special case: Linear GMRFs or linear Gaussian process.

Spatial Binary Data

For now, we focus on spatial binary data on a lattice. Nice starting point because:

- Very common in several fields. e.g. many ecological data sets
- Gaussian Markov random fields or Gaussian processes are clearly inappropriate when the data are binary. In contrast, GMRFs, GPs are often still reasonable for spatial count data.
- Can discuss both automodels (autologistic) and SGLMMs (logistic/probit) in this context.

Uncentered Autologistic Model (Besag, 1974)

If we let $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ be the random field of interest with Z_i the observation at the ith lattice point. The full conditionals:

$$\log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = \mathbf{X}_i \boldsymbol{\beta} + \sum_{j \neq i} \eta_{ij} Z_j, \tag{1}$$

where $\mathbf{X}_i$ is the ith row of the design matrix, $\boldsymbol{\beta}$ are the regression parameters, and $\boldsymbol{\eta} = \{\eta_{ij}\}$ are dependence parameters such that $\eta_{ij} \neq 0$ iff Z_i and Z_j are neighbors.

The joint distribution is given by

$$\pi(\mathbf{Z} \,|\, \boldsymbol{\theta}) = c(\boldsymbol{\theta})^{-1} \exp\left(\sum_i Z_i \mathbf{X}_i \mathbf{s} \beta + \frac{1}{2} \sum_{i,j} \eta_{ij} Z_i Z_j \right), \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')'$ and $c(\boldsymbol{\theta})$ *is the intractable normalizing constant.*

Centered Autologistic Model

The full conditional distributions for the centered model:

$$\log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = \mathbf{X}_i\boldsymbol{\beta} + \sum_{j \neq i} \eta_{ij}(Z_j - \mu_j),$$

where μ_j is the independence expectation of Z_j:

$$\mu_j = \mathbb{E}(Z_j \,|\, \boldsymbol{\eta} = \mathbf{0}) = \frac{\exp(\mathbf{X}_j\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_j\boldsymbol{\beta})}.$$

See Kaiser and Caragea (2009) for details.

Hughes, Haran and Caragea (2010) derive the joint distribution for the centered autologistic model and the corresponding algorithms for perfect sampling, maximum likelihood and Bayesian inference, utilizing parallel computing whenever possible, e.g. parametric bootstrap.

Traditional Autologistic Model: Issues

(1) **Spatial confounding**: this can be alleviated by centering.

- The autocovariate in the centered model is $Z_j - \mu_j$. This is the residual after fitting the "large scale structure" = the mean μ_j under spatial independence. $\mu_j = E(Z_j \mid \beta, \eta = 0)$
- The autocovariate in the uncentered model, Z_j, leads to uninterpretable parameters because the autocovariate includes both large scale and "small scale" structure.

(2) **Computation**: due to intractable normalizing function.

- Markov chain Maximum Likelihood (Geyer and Thompson, 1992) and an auxiliary MCMC approach using perfect sampling (Møller et al., 2006) solve the problem in the case of ML and Bayesian inference respectively.
- Maximum pseudolikelihood (MPLE) + parametric bootstrap is simpler and works well for reasonably large lattices.

Conclusions About Autologistic Models

- Always use centered autologistic model.
- For small lattices ($n < 900$), too little information to infer η unless η is large, and too little information to reliably infer $\boldsymbol{\beta}$ irrespective of the degree of dependence.
- For large lattices, inference for $\boldsymbol{\beta}$ is reliable so long as η is not too large, and inference for η is reliable when the strength of dependence is at least moderate.
- MPLE, MCMLE, and Bayes perform comparably for larger lattices. MPLE: easy to implement and fast.
- MCMC mixes better for centered model ($\eta, \boldsymbol{\beta}$ "de-correlated.")

See Hughes, Haran, Caragea (2010).

Another Framework for Non-Gaussian Data: SGLMMs

- There is a need for more flexibility and a broader framework that will easily allow for other kinds of non-Gaussian data.
- Automodels may have undesirable properties. e.g. non-intuitive implied dependencies (Wall, 2004; Assuncao and Krainski, 2009.)
- Spatial Generalized Linear Mixed Models (SGLMMs) are a much broader class of models, well suited to a variety of Gaussian and non-Gaussian spatial data.

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:

$$\mathbf{Z}|\Theta, \beta \sim N(\mathbf{X}\beta, Q^{-1}(\Theta))$$

- Note: for some popular forms of the Gaussian Markov random field the precision matrix is singular. Still allows for Bayesian inference.
- For continuous-domain spatial data: replace GMRF with a Gaussian process.

Spatial Generalized Linear Mixed Models

- Stage 1: Model $Z(\mathbf{s}_i)$ conditionally independent with distribution f given parameters β, Θ, spatial errors $W(\mathbf{s}_i)$

$$f(Z(\mathbf{s}_i)|\beta, \Theta, W(\mathbf{s}_i)),$$

  where $g(E(Z(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = X(\mathbf{s}_i)\beta + W(\mathbf{s}_i)$, η is a canonical link function (for example the logit link).
- Stage 2: $\mathbf{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n))^T$. Model $\mathbf{W}$ as spatially dependent either via a Gaussian Markov random field or Gaussian process.
- Stage 3: Priors for Θ, β.
- Inference based on $\pi(\Theta, \beta, \mathbf{W} \mid \mathbf{Z})$.

Originally Besag et al. (1991), Diggle et al. (1998), but also see Banerjee et al. (2004), Rue and Held (2005), Haran (2010), ...

SGLMMs: Challenges

SGLMMs are very flexible and have become very popular even outside mainstream statistics. They do, however, pose two major challenges:

(1) Confounding between spatial random effects and covariates (fixed effects).

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

(2) Computational challenges due to high dimensional spatial random effects. Two-pronged issue:

- MCMC is slow per iteration due to high dimensionality.
- Resulting Markov chain is slow mixing due to dependence.

Rue and Held (2002), Christensen et al. (2006), Haran, Hodges, Carlin (2003), Haran and Tierney (2010)

Spatial Confounding in SGLMMs

Introducing spatial random effects may lead to biased estimate of the posterior distribution of β and an incorrect estimate of its variance. Why this happens:

- Let $\mathbf{P}$ be orthogonal projection onto $\text{span}(\mathbf{X})$, $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
- Let $\mathbf{P}^{\perp}$ be orthogonal projection onto $\text{span}(\mathbf{X})$'s orthogonal complement, $\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P}$.
- Spectral decomposition to acquire orthogonal bases, $\mathbf{K}_{n\times p}$ and $\mathbf{L}_{n\times(n-p)}$, for $\text{span}(\mathbf{X})$ and $\text{span}(\mathbf{X})^{\perp}$, respectively. These bases allow us to write:

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

which exposes the source of the spatial confounding: $\mathbf{K}$ is collinear with $\mathbf{X}$.

Spatial Confounding: Reparameterization Solution

- Reich, Hodges and Zadnik (2006) propose solution: since **K** have no scientific meaning, delete them from the model.
- RHZ model:

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

And the prior for the random effects, δ, is now

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

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

The RHZ reparameterization corrects bias and variance issues.

Slight reduction in number of parameters: $n + p + 1$ to $n + 1$.

This reparameterization ignores the underlying graph (lattice).

A Sparse Reparameterization

- Represent the graph/lattice $G = (V, E)$ using its adjacency matrix, $\mathbf{A}$, which is the $n \times n$ matrix with entries given by $\text{diag}(\mathbf{A}) = \mathbf{0}$ and $\mathbf{A}_{ij} = 1\{(i,j) \in E, i \neq j\}$, where $1\{\cdot\}$ denotes the 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),

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

A Sparse Reparameterization [cont'd]

- 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 $\mathbf{P}^{\perp}$.
- The resulting operator, $\mathbf{M}_{\mathbf{X}}(\mathbf{A}) = \mathbf{P}^{\perp}\mathbf{A}\mathbf{P}^{\perp}$, our Moran operator for $\mathbf{X}$ with respect to the graph G, appears in the numerator of a generalized form of Moran's I:

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

- The eigenvectors comprise all possible patterns of clustering residual to $\mathbf{X}$ and accounting for G. The positive (negative) eigenvalues correspond to varying degrees of positive (negative) spatial dependence. (Boots and Tiefelsdorf, 2000)

A Sparse Reparameterization [cont'd]

- Replacing **L** with **M** in the RHZ model gives

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

  And the prior for the random effects is now

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

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

- Corrects bias and variance issues.
- Number of parameters reduced from $n + p + 1$ to $q + p + 1$. Note that q is $n/4$ in the examples that follow but could be much smaller depending on sparsity of the model.

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 estimates for Sparse and RHZ approaches are quite similar. Point estimate for Traditional is very poor.
- Variance estimates for Sparse and RHZ are similar but Traditional is very poor: it includes 0.
- Simulation results appear to be quite similar for other SGLMMs including for Poisson and Gaussian (linear).

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 and mixes faster.
- Can potentially obtain greater speed-ups by further reducing dimensionality if there is enough sparsity.

Generalizations and Ongoing Work

- These ideas apply quite naturally to other SGLMMs. So far we have obtained promising results for count data (Poisson model) and the linear spatial model.
- Fitted the different models to a real data set.
- These ideas can, in principle, also be applied to SGLMMs with underlying GPs.
- This framework allows for the use of other measures of spatial dependence, for e.g. one based on Geary's C instead of Moran's I.
- This is a dimension reduction approach that reduces the number of latent variables. (Several others, e.g. fixed rank kriging, kernel convolutions, Gaussian predictive process.)

Summary

- SGLMMs provide a very general approach for modeling non-Gaussian spatial data.
- Regression coefficients for such models are often uninterpretable. Our reparameterization results in interpretable regression coefficients.
- Computation can become prohibitive for SGLMMs (as n gets large.) Our approach greatly reduces the number of spatial random effects, allowing for dramatic speed-ups in computational efficiency, both computing time and mixing of Markov chain. This allows potentially for the analysis of much larger data sets than typically feasible.