Spatial Random Effects Models
A Few Thoughts

Murali Haran

partially based on joint work with John Hughes

Department of Statistics, Penn State University

Operations Research

Penn State University. April 2013

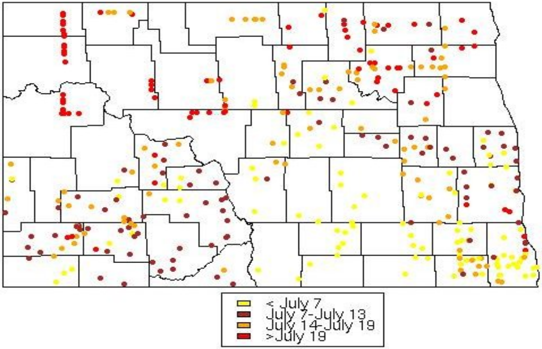
What This Talk is About

1. I will discuss the problem of specifying a model for a dependent process. The process lives in dimensions ≥ 2.
   - Want a convenient framework for both Gaussian and non-Gaussian data.
   - I will describe a convenient, flexible class of random effects models for spatial data. ("New" way of thinking about random effects.)
2. I will discuss above dependent processes as an approach for modeling smooth curves. ("Nonparametrics".)
3. Implications, research questions

This is not a talk about completed research; this is a discussion of *very basic* ideas and challenges.

Continuous Domain Spatial Data

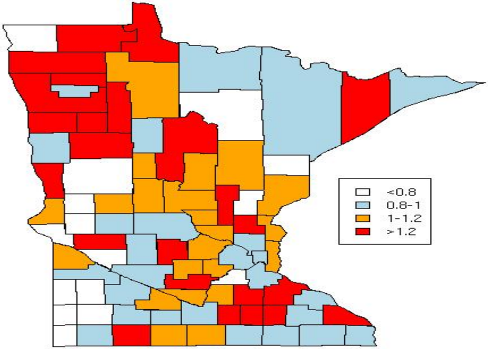
Example: wheat flowering in North Dakota (related to blight epidemics) Courtesy Plant Pathology, PSU, North Dakota State.



Haran, Bhat, Molineros, DeWolf (*J of Ag Biol Env Stats*, 2008)

Spatial Data on a Lattice

Minnesota cancer rates by county: $\frac{\text{observed}}{\text{expected}}$ counts



Courtesy MN Cancer Surveillance System, Dept. of Health

Haran, Hodges, Carlin (*J of Comp Graphical Stats*)

Forms of Spatial Data

Interested in specifying a model for an indexed stochastic process: $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, where D is spatial domain.

- "Geostatistics": $D \subset \mathbb{R}^d$. Process is infinite-dimensional
- "Areal/lattice data": D is a finite set of locations in $\mathbb{R}^d$. Data observed on or aggregated up to arbitrary spatial units.

Linear Model with Spatial Random Effects

Linear Spatial Models. Spatial version of linear regression models (though much more flexible).

- Spatial process, $Z(\mathbf{s}) = X(\mathbf{s})\beta + w(\mathbf{s}),\ \mathbf{s} \in D$, $X(\mathbf{s})$ are covariates at $\mathbf{s}$.
- Model dependence among spatial random variables by imposing it on random effects $\{w(\mathbf{s}), \mathbf{s} \in D\}$.
- Model for random effects:
  - **Gaussian process** (GP) for continous-domain.
  - **Gaussian Markov random field** (GMRF) for lattice data.

Spatial Linear Model with GP Random Effects

- Gaussian Process (GP): Infinite-dimensional stochastic process. Process at any finite collection of locations is multivariate normal.
- Dependence modeled via covariance function. Random variates closer to each other are more dependent. E.g. $\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/\phi), \phi > 0$.
- Let Θ be covariance function parameters, so covariance matrix is $\Sigma(\Theta)$.
- Let $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$

$$\mathbf{Z}|\Theta, \boldsymbol{\beta} \sim N(\mathbf{X}\boldsymbol{\beta}, \Sigma(\Theta))$$

Spatial Linear Model with GMRF Random Effects

- Gaussian Markov Random field (GMRF): Let Θ be the parameters for precision matrix $Q(\Theta)$

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

$$f(\mathbf{Z}|\Theta, \boldsymbol{\beta}) \propto c(\Theta) \exp\left(-\frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T Q(\Theta)(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right)$$

- Think about modeling in terms of conditional specifications:

$$Z(\mathbf{s}_i) \mid \{Z(\mathbf{s}_j), \text{ for } j \neq i\} \sim N\left(\frac{\sum_{j:j\sim i} Z(\mathbf{s}_j)}{n_i}, \frac{\sigma^2}{n_i}\right),$$

where $j \sim i$ if $Z(\mathbf{s}_j)$ and $Z(\mathbf{s}_i)$ are neighbors, n_i is the number of neighbors for ith node on lattice.

Purpose of Fitting These Models

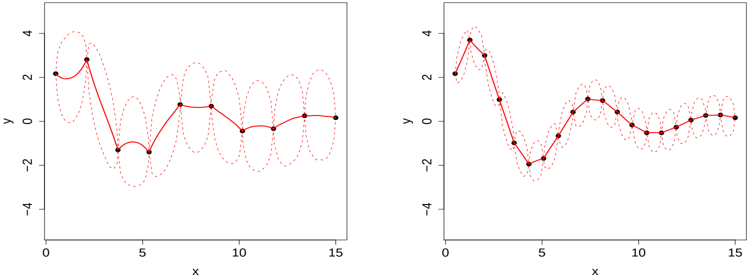
- Continuous-domain model: model dependence appropriately. Principled interpolation – accounting for uncertainties when interpolating.
- Discrete (lattice) model: model dependence appropriately. Spatial smoothing.

So far: Described random effects models for dependent data. However, **these random effects models are also very useful as flexible models for smooth curves**.

Function Interpolation

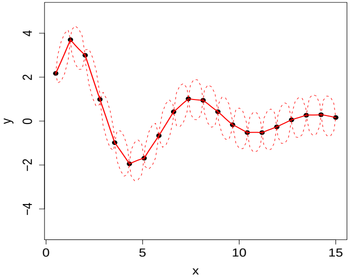
Interpolations using simple GP random effects model:

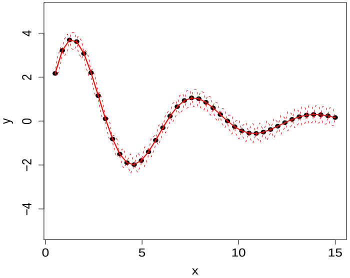
$y(x) = \mu + w(x)$, $\{w(x), x \in (0, 20)\}$ is a zero-mean GP.



Increase data from 10 to 20 points

Function Interpolation





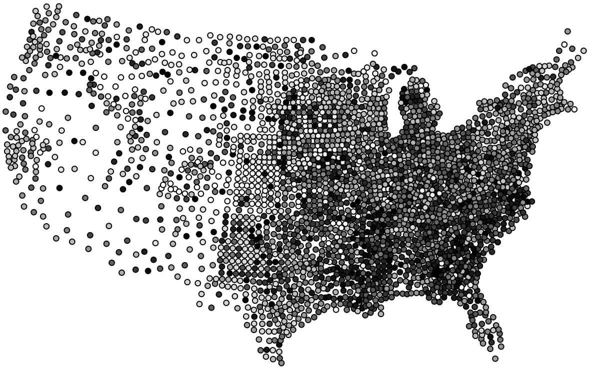
Increase data from 20 to 40 points

Random Effects and Penalized Splines

- Direct correspondence in some cases between fitting these models and fitting splines (cf. Grace Wahba's work).
- Consider fitting a penalized spline to data. For certain splines, minimizing the objective function is formally identical to estimating a simple random effects (mixed linear) model (Ruppert, Wand, Carroll, 2003).

Non-Gaussian Spatial Data

- How about models for non-Gaussian dependent data? For e.g. spatial binary data or count data? Very common in many disciplines
- Figure: U.S. infant mortality data by county. $n = 3071$



Non-Gaussian Spatial Data

- Continuous-domain: tough to specify a joint model directly, with covariance decaying with distance. Easy to specify: Gaussian process.
- Specify joint model via full conditionals. Non-trivial. For e.g. for lattices need Hammersley-Clifford Theorem and/or Brook's Lemma. Easy to specify: Gaussian Markov random field.
- One nice solution: Use generalized linear model framework with underlying Gaussian model. This allows us to continue to use the GP or GMRF model specifications to build models for non-Gaussian data.

Spatial Generalized Linear Mixed Models (SGLMMs)

- Stage 1: Model $Z(\mathbf{s}_i)$ conditionally independent with distribution f given parameters $\boldsymbol{\beta}, \Theta$, spatial errors $w(\mathbf{s}_i)$

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

  where $g(E(Z(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = X(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$, η is a canonical link function.
- Stage 2: Again $\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))^T \sim$ GP or GMRF.

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

Spatial Generalized Linear Mixed Models

- Lattice model: $G = (V, E)$ is underlying graph
- Stage 1: model for spatial data Z at location $\mathbf{s}_i$
  - $f(Z(\mathbf{s}_i)|\boldsymbol{\beta}, \Theta, W(\mathbf{s}_i)), i = 1, \ldots, n$, conditionally independent
    E.g. $Z(\mathbf{s}_i) \mid \boldsymbol{\beta}, W(\mathbf{s}_i) \sim$Poisson($\mu_i$)
  - $g(E(Z(\mathbf{s}_i))) = X(\mathbf{s}_i)\boldsymbol{\beta} + W(\mathbf{s}_i)$
    E.g. $\log(\mu_i) = X(\mathbf{s}_i)\boldsymbol{\beta} + W(\mathbf{s}_i)$
- Stage 2: $\mathbf{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n))^T$
  W on a lattice: Gaussian Markov random field

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

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

SGLMMs

Like linear random effects model, now useful for non-linear regression in a generalized linear model framework. SGLMMs useful as:

- Models for spatial data
- Machine learning, nonparametric regression, classification

SGLMM Inference and Prediction

SGLMMs have become very popular even outside mainstream statistics.

(1) Prediction: very powerful and flexible.

(2) Inference:

- Interpretation: flexibility poses challenges for interpretability of regression coefficients, similar to confounding issues in standard regression.
- Computation: with many data points, high-dimensional spatial random effects pose considerable challenges.

Inference for Linear Spatial Models

Why computing may be challenging:

- Linear model: likelihood $\mathcal{L}(\Theta, \beta; \mathbf{Z})$ looks like $f(\mathbf{Z}|\Theta, \beta) \propto c(\Theta) \exp\left(-\frac{1}{2}(\mathbf{Z} - \mathbf{X}\beta)^T Q(\Theta)(\mathbf{Z} - \mathbf{X}\beta)\right)$
- Can perform ML estimation or Bayesian inference.
- When $\mathbf{Z}$ is high-dimensional, the matrices are high-dimensional. Not a problem in the case of GMRFs due to sparse precision matrix. Problematic with GP due to dense covariance.

Inference for SGLMMs

Why computing may be more challenging:

- SGLMM: Unlike in linear case cannot "marginalize out" spatial random effects **w** ahead of time (in linear case both were normal models so this was easy).
- ML inference: maximize *marginal* likelihood – need to integrate out high-dimensional spatial random effects. Challenging problem.
- More convenient to carry out Bayesian inference. Specify priors on parameters so inference based on posterior: $\pi(\Theta, \boldsymbol{\beta}, \mathbf{w}|\mathbf{Z}) \propto \mathcal{L}(\mathbf{Z}|\Theta, \boldsymbol{\beta}, \mathbf{w}) f_{\mathbf{w}}(\mathbf{w}|\Theta) f_{\Theta}(\Theta) f_{\boldsymbol{\beta}}(\boldsymbol{\beta})$
- Need Markov chain Monte Carlo. Challenging:
  - Slow mixing Markov chain.
  - Each update of Markov chain may be expensive.

Confounding Issues

- In simple regression $Y = X_1\beta_1 X_2\beta_2 + \epsilon$, if there is multi-collinearity between X_1 and X_2, estimates of β_1, β_2 are unreliable.
- Similarly, in spatial linear model, random effect w may be collinear with the "regular" covariate X. Hence, the flexibility of the spatial random effect can make β inference/interpretation very unreliable.
- Solution: Reich, Hodges, Zadnik (2006) and Hughes, Haran (2012): rewrite the model to tease apart random effects that are collinear from those that are not. Delete the collinear random effects. The resulting model is more interpretable and much easier to fit.

Using These Models

- Main observation: here random effects are a device for inducing dependence, adding flexibility. We have more flexibility when working with them. This is basis for solving confounding issue.
- This also is the basis for thinking about improved modeling/computational strategies.
- This should also influence how we think about these models. For e.g. simulation studies – use the functional form of interest to simulate rather than simply simulating from the random effects model. The random effects model may not be of interest in its own right.

Summary

- Spatial random effects models are a very general, flexible approach for modeling dependent processes, both Gaussian and non-Gaussian.
- These models are also very flexible and useful for non-linear regression, both Gaussian and non-Gaussian.
- Challenges: Computation, confounding issues.
- Important to think about how they are devices for producing flexible models; this should influence modeling approaches as well as how we think about these models overall, e.g. how we design simulation studies.

Spatial Confounding in SGLMMs

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

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

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

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

Spatial Confounding: Reparameterization Solution

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

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

  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.
- RHZ approach ignores underlying graph

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

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

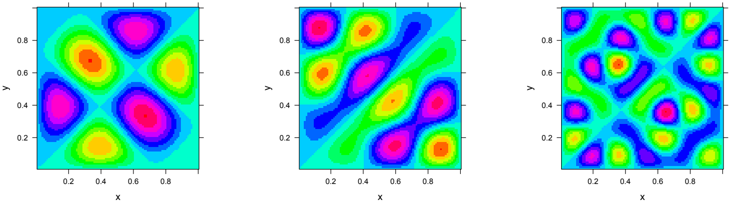
  where $\mathbf{Q}^{**} = \mathbf{M}'\mathbf{Q}\mathbf{M}$.

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

Interpreting the Resulting Reparameterization

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

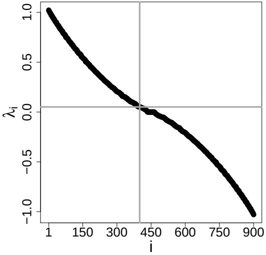
Some selected basis vectors for the 30×30 lattice.



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.



Study: Inference for Spatial Binary

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

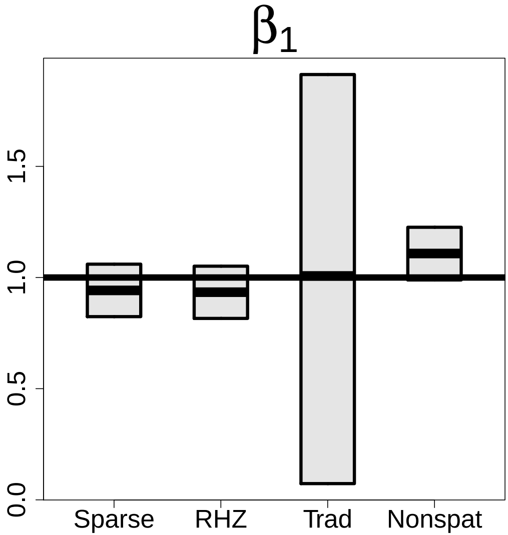
Predictors are the coordinates of unit square.

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 Poisson and Gaussian (linear)

Spatial Count Data: Simulation Results



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
- MCMC easier to construct. Computational efficiency allows for more 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. (2012) "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models," *Journal of the Royal Statistical Society (B), in press.*

Software: http://www.biostat.umn.edu/∼johnh/software.html