Inference for Count Data using the Spatial Random Effects Model

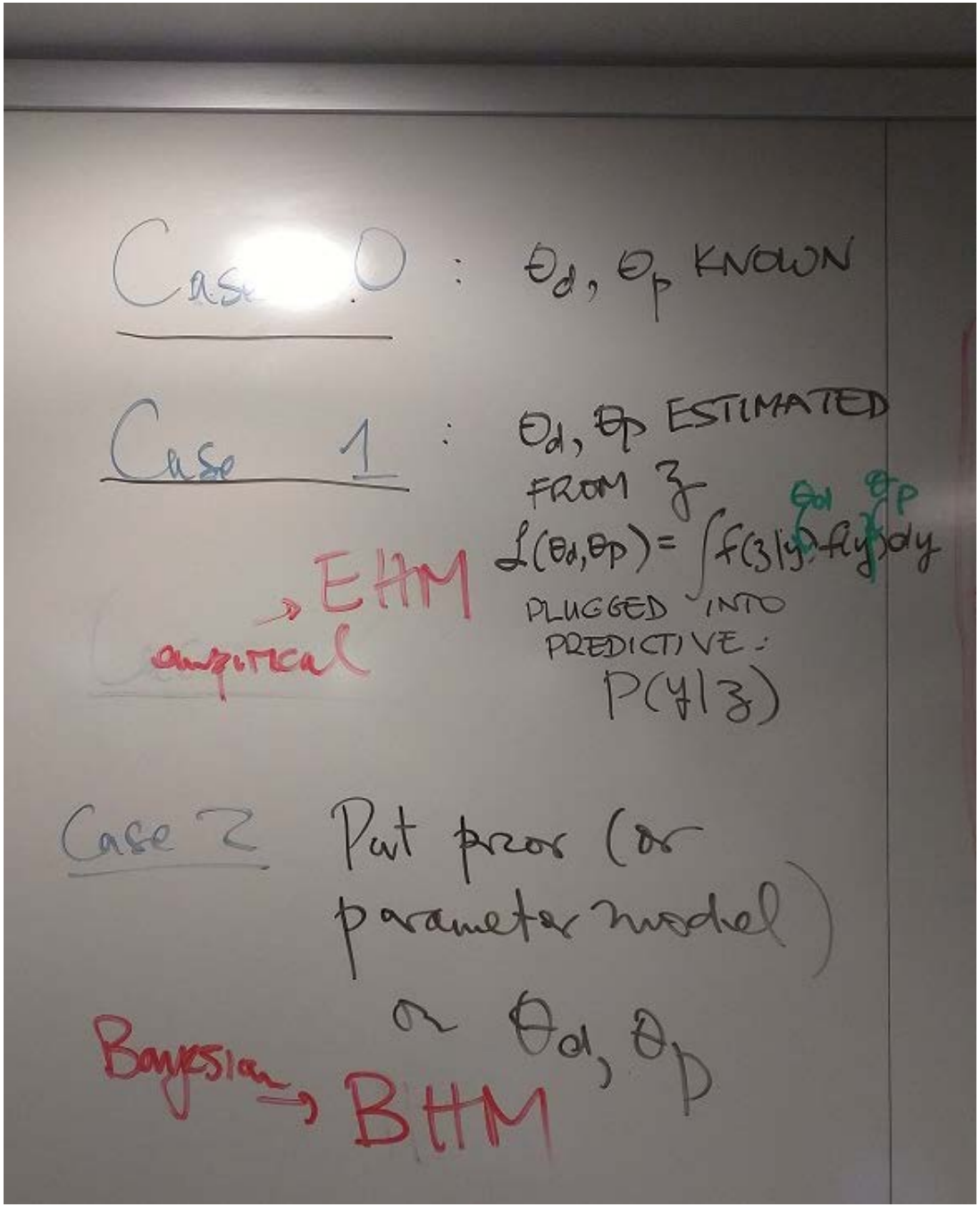
Noel Cressie

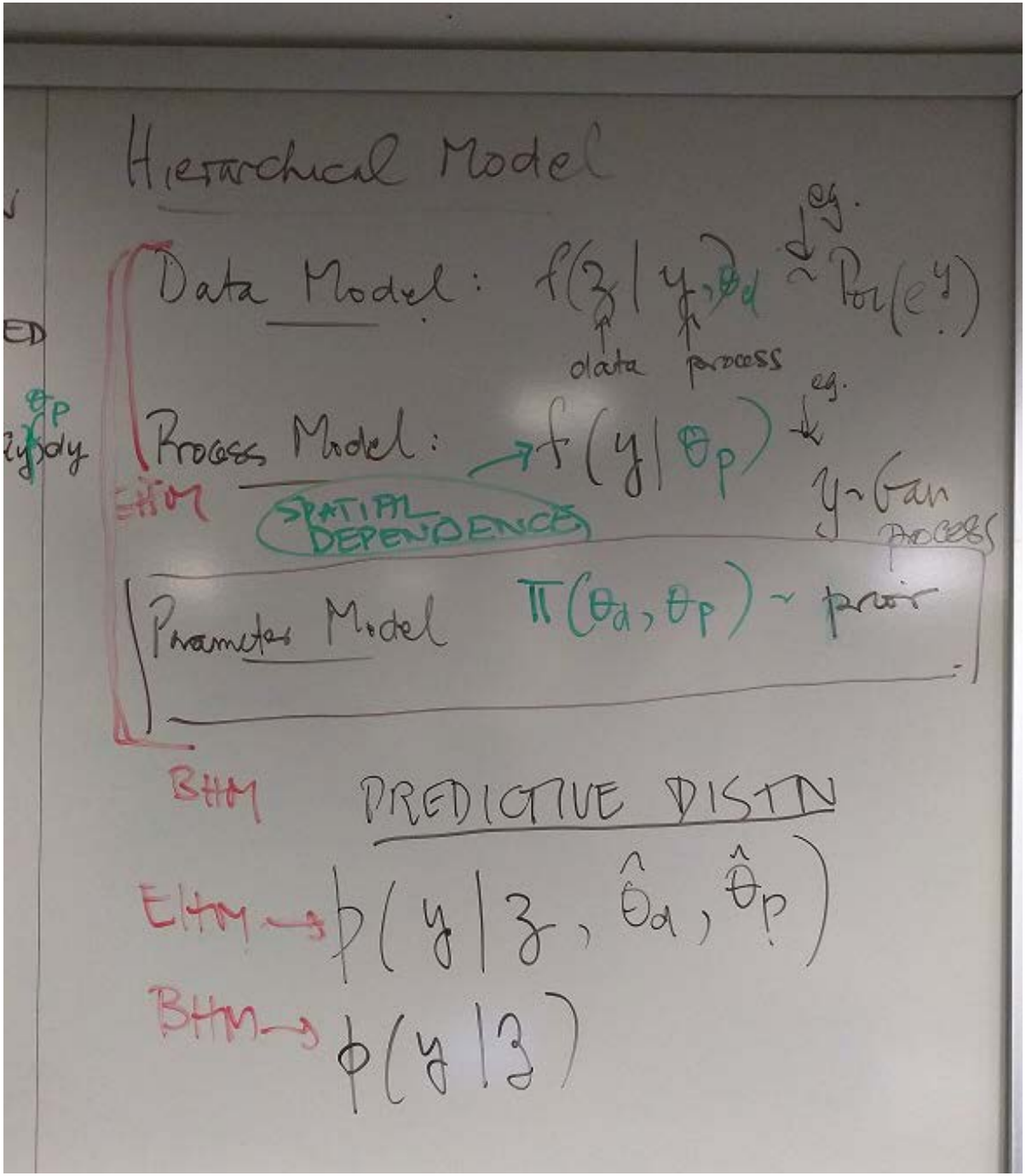
National Institute for Applied Statistics Research Australia (NIASRA)

University of Wollongong

(ncressie@uow.edu.au)

Joint research with **Aritra Sengupta**


Case 0 : θ_d, θ_p KNOWN
Case 1 : θ_d, θ_p ESTIMATED FROM $\mathcal{Z}$
$\mathcal{L}(\theta_d, \theta_p) = \int f(\mathcal{Z}|y, \theta_d) f(y|\theta_p) dy$
PLUGGED INTO PREDICTIVE: $P(y|\mathcal{Z})$
→ EHM
analytical
Case 2 Put prior (or parameter model) on θ_d, θ_p
Bayesian → BHM


Hierarchical Model
Data Model: $f(z \mid y, \theta_d)$ eg. $\sim Pois(e^y)$
data
process
Process Model: $f(y \mid \theta_p)$ eg. $y \sim$ Gau process
SPATIAL DEPENDENCE
Parameter Model $\pi(\theta_d, \theta_p) \sim$ prior
EHM
BHM
PREDICTIVE DISTN
EHM → $\phi(y \mid z, \hat{\theta}_d, \hat{\theta}_p)$
BHM → $\phi(y \mid z)$

$$p(y \mid z, \hat{\theta}_d, \hat{\theta}_p)$$

$$\propto f(z, y \mid \hat{\theta}_d, \hat{\theta}_p) \quad \text{DATA MODEL}$$

$$= \frac{f(z \mid y, \hat{\theta}_d) f(y \mid \hat{\theta}_p)}{\int f(z, t, \hat{\theta}_d) f(t, \hat{\theta}_p)\, dt} \quad \text{PROCESS MODEL}$$

NORMALIZING (MARGINAL $f(z)$)

$\Rightarrow$ MCMC!

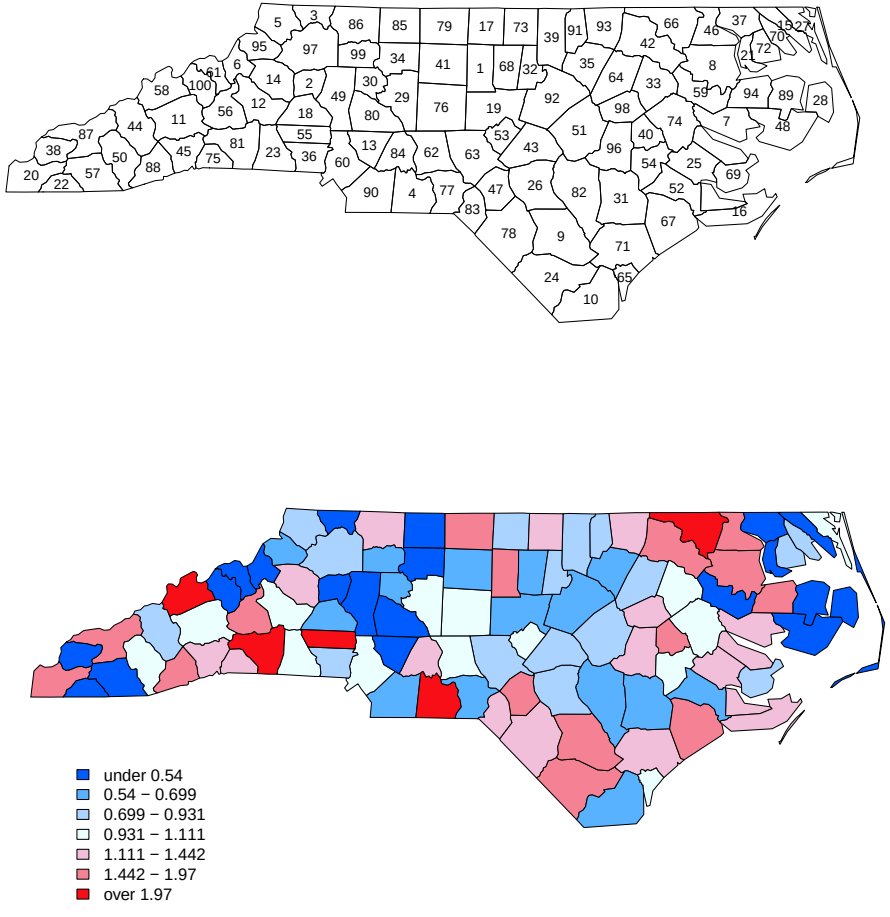
EVEN FOR NON-BAYESIAN

(2016) Bradley, Wikle, & Holan JASA

Areal Data Consisting of Counts: Examples

- Numbers of a given invasive species distributed over small areas.
- Health studies : Number (or rates) of individuals with a disease.
- Remote sensing of the environment : Number of pixels within a small region having some desired property (e.g., number of cloudy pixels).
- Socio-economic studies : Unemployment rates by counties.

An Example: Count Data Over the Counties of NC



Chloropleth map of counties of NC showing the standardized mortality ratio (SMR) of SIDS, namely the ratio of (observed # SIDs)/(expected # SIDs).

Common Modeling Approaches for Count Data

- Modeling approaches based on a Markov random field (e.g., Besag *et al.* [1991], Cressie [1993], Rue and Held [2005]).
- Poisson kriging (e.g., Monestiez *et al.* [2006])
- Approximate Gaussian models (e.g., Cressie and Chan [1989])
- Fully Bayesian approach for geostatistical models based on stationarity assumptions (e.g., Diggle *et al.* [1998], Zhang [2002]).

Our Hierarchical Modeling Strategy: Count data is modeled through a Generalized Linear Mixed Model (GLMM) (e.g., Diggle *et al.* [1998]), where the data, conditional on an **underlying spatial process**, has a Poisson distribution. The hidden underlying spatial process is modeled using a transformed **spatial Gaussian Field (GF)** whose means, variances, and covariances depend on unknown parameters. We **estimate** the parameters and **predict** the hidden spatial process. (The Poisson assumption can be expanded to one involving a zero-inflated Poisson; see Sengupta *et al.* [2012].)

Hierarchical Statistical Modeling (HM)

Step away from remote sensing for the moment to get a broader perspective on uncertainty quantification in science.

- Let Y be the data, X be the process (or state) of interest, and θ be unknown parameters. (For example, in my research in spatio-temporal statistics, Y might have dimension $10^6 - 10^9$, X might be of the same order, and θ might have dimension $10^2 - 10^4$.)
- [A|B] denotes the conditional distribution of generic quantity A, given generic quantity B; and [B] denotes the distribution of B.
- [A,B] denotes the joint distribution of A and B. Then

$$[\text{A,B}] = [\text{A}|\text{B}] \cdot [\text{B}].$$

HM Captures Sources of Uncertainty

Sources of uncertainty: the data, the process, and the parameters.

- All uncertainties can be expressed through the joint distribution, $[Y, X, \theta]$. From the previous slide,

$$\begin{aligned}[Y, X, \theta] &= [Y, X|\theta] \cdot [\theta] \\ &= [Y|X, \theta] \cdot [X|\theta] \cdot [\theta]\end{aligned}$$

- Data model: $[Y|X, \theta]$
- Process model: $[X|\theta]$
- Parameter model: $[\theta]$
- This implies an additive decomposition of the joint entropy: $E(\log [Y, X, \theta]) = E(\log [Y|X, \theta]) + E(\log [X|\theta]) + E(\log [\theta])$.

Predictive Distribution

- Three cases:
  0. θ is known, and hence the parameter model is degenerate at θ. The posterior distribution is $[X|Y,\theta]$ – since θ is known, this is also the predictive distribution.
  1. θ is fixed but unknown, and it is estimated from the data Y; call the estimate $\widehat{\theta}$. The parameter model is assumed degenerate at $\widehat{\theta}$ and the (empirical) predictive distribution is $[X|Y,\widehat{\theta}]$.
  2. θ is unknown, and its uncertainty is captured with the parameter model $[\theta]$. The posterior distribution is $[X,\theta|Y]$, and the predictive distribution is $[X|Y]$.
- Case 0. is often unrealistic; case 1. is called empirical hierarchical modeling (EHM); and case 2. is called Bayesian hierarchical modeling (BHM).

Inference on the State X

- X and θ appear asymmetrically: there is usually more interest in inference on X.
- Since X is uncertain, it is modeled as a random quantity. Inference on a random quantity is sometimes called "prediction." Hence, we use the term predictive distribution; for the three cases, it is:

  Case 0. $[X|Y,\theta] = [Y|X,\theta] \cdot [X|\theta]/[Y|\theta]$

  Case 1. $[X|Y,\widehat{\theta}] = [Y|X,\widehat{\theta}] \cdot [X|\widehat{\theta}]/[Y|\widehat{\theta}]$,

  Case 2. $[X|Y] = \int [Y|X,\theta] \cdot [X|\theta] \cdot [\theta] d\theta / [Y]$,

  and it is not always the same as the posterior distribution.
- We still say "estimation of X," rather than prediction of X.
- Inference on X should be based on the predictive distribution. This is fundamental to Uncertainty Quantification!

Computational Issues

- Hierarchical modeling approaches that are based on a full-rank model for the underlying spatial process, are not computationally feasible when the data size is large.
- Reduced-rank modeling approaches for the underlying spatial GF have been developed to deal with this computational challenge (e.g., Wikle [2002], Cressie and Johannesson [2006, 2008], Banerjee *et al.* [2008], Stein [2008], Lopes *et al.* [2008, 2011]).
- Many of the spatial and spatio-temporal applications for very-large-to-massive datasets center around reduced-rank representations of geostatistical models, where the hidden process is modeled using a continuous GF (e.g., see the review in Wikle [2010]).
- Lindgren *et al.* [2011] show how the underlying spatial process can be modeled using a Gaussian Markov Random Field (GMRF) in a computationally efficient manner.

Spatial Statistical Modeling Strategy

- **Classical spatial statistical models:**
  - Non hierarchical models based on a MRF often give rise to complicated likelihoods and computational bottlenecks.
  - Spatial outliers can be difficult to deal with in a nonhierarchical model.
  - Stationarity or homogeneity assumptions are not valid in a lot of situations (e.g., ecological modeling of an invasive species).
- **Parameters:** We want to capture sources of variability but avoid putting inappropriate priors on parameters. This is what a classical geostatistical analysis does.
- **We propose:**
  - a data model based on the Poisson distribution;
  - a spatial statistical process model based on a (log) Gaussian geostatistical SRE model (Cressie and Johannesson [2006, 2008]);
  - parameters that are fixed but unknown (and hence will be estimated).

Model Specification for Areal Count Data

- Let the spatially related areas be $\{A_i : 1 = 1, \ldots, N\}$ located at $\{\mathbf{s}_i : i = 1, \ldots, N\}$. The ordering of the indices is unimportant.
- The spatial domain of interest is $\bigcup_{i=1}^{N} A_i$; define the spatial index set $D \equiv \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$.
- Some (or all) of the areas have counts, $\{Z(\mathbf{s}_i) : i = 1, \ldots, n\}$, associated with them, where $n \leq N$. Hence $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_N\}$ represents areas $\{A_{n+1}, \ldots, A_N\}$ where there are no data.
- The observations are $\mathbf{Z}_O \equiv (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^\top$.

Model Specification for Areal Count Data [cont.]

- **Data model:** Counts, $Z(\mathbf{s}_i)|\mu_{Z|Y}(\mathbf{s}_i) \sim$ ind. Poisson$(\mu_{Z|Y}(\mathbf{s}_i))$; $i = 1, \ldots n$. The set $\{\mathbf{s}_i : i = 1, \ldots, n\}$ is the set of locations of small areas where there are data. The mean count $\mu_{Z|Y}(\mathbf{s}_i)$ is modeled for all $i = 1, \ldots, N$. Note that the data model can be generalized to the zero-inflated Poisson.
- **Link function:**

$$Y(\mathbf{s}_i) = \log(\mu_{Z|Y}(\mathbf{s}_i)),$$

  where $\log(\cdot)$ is the link function. Note that other link functions are possible.
- **Process model:**

$$\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}}).$$

  - $\mathbf{Y} \equiv (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_N))^\top$, where $N \geq n$.
  - The set $\{\mathbf{s}_i : i = 1, \ldots, N\}$ is the set of all locations. Recall that the locations $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_N\}$ represent small areas where there are no data.
  - The process model captures spatial dependencies through $\boldsymbol{\Sigma}_{\mathbf{Y}}$ (geostatistical model); cf. capturing spatial dependence using $\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}$ (GMRF model or spatial econometric model).

SRE Model: A Component of the Hidden Process $Y(\cdot)$

- Model for $Y(\cdot)$: $Y(\cdot) = C(\cdot) + \mathbf{x}(\cdot)^\top \boldsymbol{\beta} + W(\cdot)\left\{\mathbf{S}(\cdot)^\top \boldsymbol{\eta} + \xi(\cdot)\right\}$
- $C(\cdot)$ is a known offset term and $W(\cdot)$ is a known weight.
- $\mathbf{x}(\cdot)^\top \boldsymbol{\beta}$ is the large-scale spatial variation.
- $\mathbf{S}(\cdot)$ is the r-dimensional vector of multiresolution spatial basis functions.
- $\boldsymbol{\eta}$ is an r-dimensional vector of random effects, where $r << n \leq N$.
- Distribution for $\boldsymbol{\eta}$: $[\boldsymbol{\eta}|\mathbf{K}] \sim \mathrm{N}(\mathbf{0}, \mathbf{K})$.
- $\xi(\cdot)$ is spatially unstructured and represents fine-scale variability:

$$\xi(\cdot) \sim \text{ind. } \mathrm{N}(0, \sigma_\xi^2 v_\xi(\cdot)), \text{ where } \sigma_\xi^2 > 0 \text{ and } v_\xi(\cdot) \text{ is known.}$$

- Covariance matrix for $\mathbf{Y}$ (Cressie and Johannesson [2008]): $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \mathbf{W}\left(\mathbf{SKS}^\top + \sigma_\xi^2 \mathbf{V}_\xi\right)\mathbf{W}$, where $\mathbf{V}_\xi \equiv \mathrm{diag}(v_\xi(\mathbf{s}_1), \ldots, v_\xi(\mathbf{s}_N))$ and $\mathbf{W}$ are $N \times N$ diagonal matrices.
- Denote: $\mathbf{V}_{\xi,O} = \mathrm{diag}(v_\xi(\mathbf{s}_1), \ldots, v_\xi(\mathbf{s}_n))$.

Goal of our Research

- Our ultimate goal is inference on $Y(\mathbf{s})$; $\mathbf{s} \in D \equiv \{\mathbf{s}_i : i = 1, \ldots, N\}$ from the count data $\mathbf{Z}_O$.
- We want to make predictions on $\mathbf{Y}$, from the predictive distribution, $[\mathbf{Y}|\mathbf{Z}_O, \boldsymbol{\beta}, \mathbf{K}, \sigma^2_\xi]$, which is conditional on $\mathbf{Z}_O$ **and** the parameters.
  - This predictive distribution is obtained using Bayes' Theorem, but usually it cannot be obtained in closed form.
  - We use MCMC samples from $[\boldsymbol{\eta}, \boldsymbol{\xi}|\mathbf{Z}_O, \boldsymbol{\beta}, \mathbf{K}, \sigma^2_\xi]$. Efficient MCMC computations are needed when n is very large.
- Now assume that parameters $\boldsymbol{\beta}$, $\mathbf{K}$, and σ^2_ξ are fixed but unknown, and we shall **estimate** them.
- The model is called an **Empirical Hierarchical Model (EHM)**, resulting in **Empirical-Bayesian inference**.

Inference for the Hierarchical Model

- We need the predictive distribution, $[\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}_O, \boldsymbol{\theta}]$, where in practice, the estimate $\hat{\boldsymbol{\theta}} \equiv \{\hat{\boldsymbol{\beta}}, \hat{\mathbf{K}}, \hat{\sigma}^2_\xi\}$, is substituted in for $\boldsymbol{\theta}$.
- Let $\boldsymbol{\xi}^\top \equiv (\boldsymbol{\xi}_O^\top, \boldsymbol{\xi}_U^\top)$, where $\boldsymbol{\xi}_O^\top = (\xi(\mathbf{s}_1), \ldots, \xi(\mathbf{s}_n))$, and $\boldsymbol{\xi}_U^\top = (\xi(\mathbf{s}_{n+1}), \ldots, \xi(\mathbf{s}_N))$.
- $\boldsymbol{\xi}_U$ is independent of $\mathbf{Z}_O, \boldsymbol{\xi}_O, \boldsymbol{\eta}$ (conditional on $\boldsymbol{\theta}$); hence,

$$[\boldsymbol{\xi}_U | \mathbf{Z}_O, \boldsymbol{\xi}_O, \boldsymbol{\eta}, \boldsymbol{\theta}] = [\boldsymbol{\xi}_U | \boldsymbol{\theta}].$$

- So, for all $\mathbf{s} \in \{n+1, \ldots, N\}$, we have:

$$E(\xi(\mathbf{s}) | \mathbf{Z}_O, \boldsymbol{\xi}_O, \boldsymbol{\eta}, \boldsymbol{\theta}) = E(\xi(\mathbf{s}) | \boldsymbol{\theta}) = 0$$

$$\text{var}(\xi(\mathbf{s}) | \mathbf{Z}_O, \boldsymbol{\xi}_O, \boldsymbol{\eta}, \boldsymbol{\theta}) = \text{var}(\xi(\mathbf{s}) | \boldsymbol{\theta}) = \sigma^2_\xi v_\xi(\mathbf{s}).$$

- Hence, we only need MCMC samples from the predictive distribution, $[\boldsymbol{\eta}, \boldsymbol{\xi}_O | \mathbf{Z}_O, \boldsymbol{\theta}]$, where recall that $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ is "plugged in."

EM Algorithm: Review

- Suppose we observe $X_{obs} \sim f(x_{obs}|\theta)$, and we are interested in finding the **MLE** of θ.
- If maximizing $L(\theta) \equiv f(x_{obs}|\theta)$ is hard, but maximizing the **complete data likelihood**, $L_c(\theta|x_{obs}, x_{mis}) \equiv f(x_{obs}, x_{mis}|\theta)$, is easy, we can use the EM algorithm to find the MLE of θ.

The EM algorithm (Dempster *et al.* [1997]): Start with $\theta^{[0]}$, and repeat the following steps until convergence:

E-step: Calculate

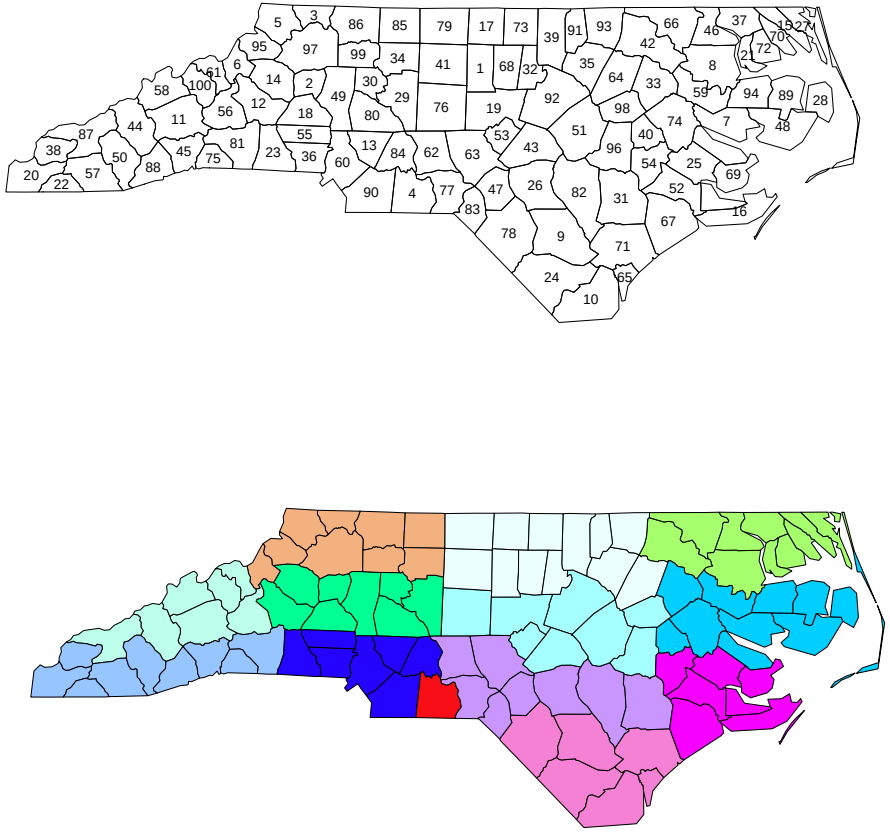
$$Q(\theta, \theta^{[l]}) \equiv E_{\theta^{[l]}} \left\{\log f(x_{obs}, X_{mis}|\theta)|x_{obs}\right\} = E_{\theta^{[l]}} \left(\log(L_c(\theta|x_{obs}, X_{mis}))|x_{obs}\right).$$

M-Step: Calculate $\theta^{[l+1]} = \arg\max_\theta Q(\theta, \theta^{[l]})$.

North Carolina SIDS Data (1974-78)

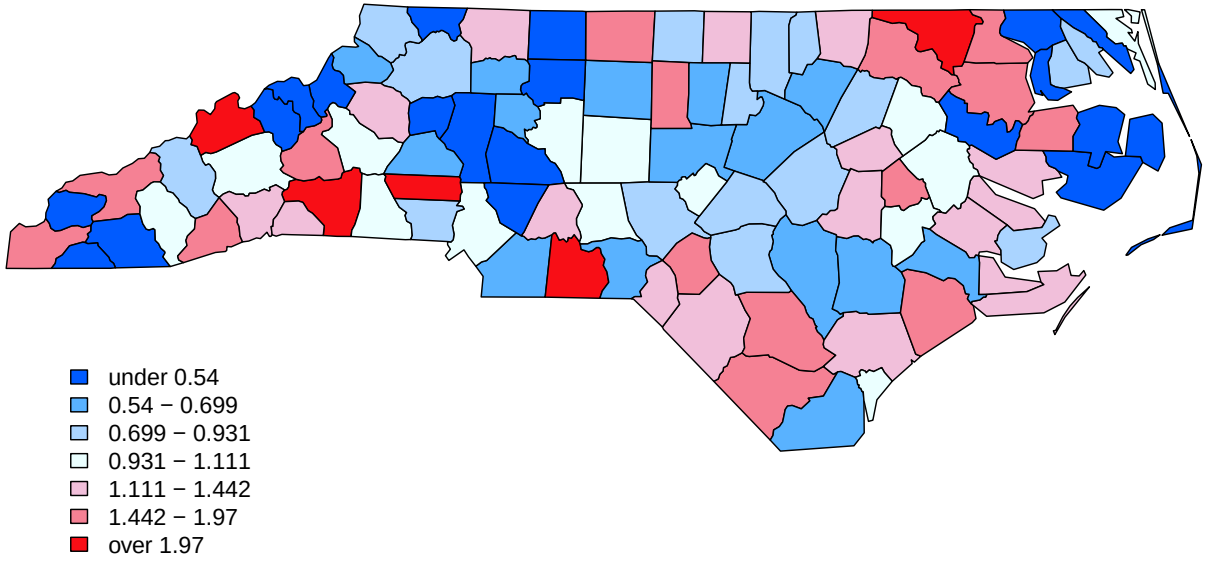
- The hierarchical model above is used to analyze the well known dataset that contains information on Sudden Infant Death Syndrome (SIDS) from 1974 to 1978, for the 100 counties of North Carolina (NC). Here $\mathbf{s}_1, \ldots, \mathbf{s}_{100}$ are the centroids of the counties, numbered alphabetically by county name.
- SIDS is a classification of death for healthy infants under one year of age.
- Cressie and Chan [1989] analyzed the NC SIDS data using a model based on the Freeman-Tukey transformation.
- We directly model SIDS for the $N = n = 100$ counties of NC using the hierarchical Poisson SRE model, where parameters are estimated. Note that Cressie and Chan [1989] omitted county 4 as an outlier (for them, n=99); we include county 4 in our analysis below.

Counties of North Carolina (NC)



County index from $1, \ldots, 100$ (top panel); and 12 contiguous regions for the counties of NC, where county 4 is shown in red (bottom panel).

Standardized Mortality Ratio (SMR) of SIDS



Chloropleth map of counties of NC showing the standardized mortality ratio (SMR) of SIDS, namely the ratio of (observed # SIDs)/(expected # SIDs).

Notation

- $Z(\mathbf{s}_i)$ denotes the observed number of SIDS in county i, $i = 1, \ldots 100$.
- $E(\mathbf{s}_i)$ is the expected number of SIDS in county i, obtained from internal standardization:

$$E(\mathbf{s}_i) = \sum_{k=1}^{4} B_k(\mathbf{s}_i) \left(\sum_{j=1}^{100} Z_k(\mathbf{s}_j) / \sum_{j=1}^{100} B_k(\mathbf{s}_j) \right), \; i = 1, \ldots 100.$$

  - $k = 1, \ldots, 4$ represents the subgroups, "white male," "white female," "non-white male," and "non-white female," respectively.
  - $Z_k(\mathbf{s}_i)$: observed number of SIDS for subgroup k in county i, $i = 1, \ldots 100$.
  - $B_k(\mathbf{s}_i)$: total number of live births for subgroup k in county i, $i = 1, \ldots 100$.
- $SMR(\mathbf{s}_i)$ is the standardized mortality ratio (SMR) in county i, and it is defined as

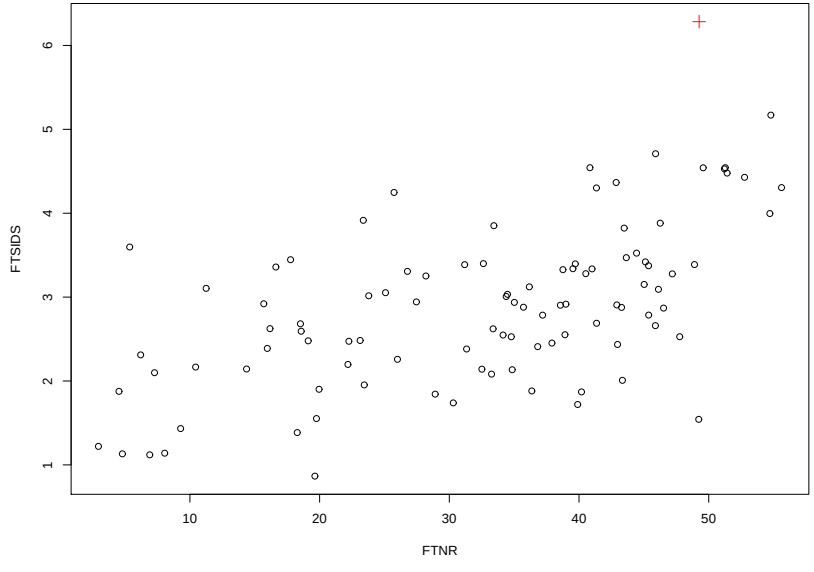
$$SMR(\mathbf{s}_i) = Z(\mathbf{s}_i)/E(\mathbf{s}_i), i = 1, \ldots 100.$$

- $SMR(\cdot)$ estimates the relative risk, $\lambda(\cdot) \equiv \mu_{Z|Y}(\cdot)/E(\cdot)$.

Exploratory Study

The Freeman-Tukey transformation of counts is essentially the square-root transformation; it removes dependence between the mean and the variance.

- $FTNR(\mathbf{s}_i)$: Freeman-Tukey transformed non-white live birth rate in county i.
- $FTSIDS(\mathbf{s}_i)$: Freeman-Tukey transformed SIDS rate in county i.
- Scatter plot of $FTSIDS$ against $FTNR$ justifies using $FTNR$ as a covariate, along with the vector of 1s. County 4 is shown as a red "+".



Standard GLM Approach

- $Z(\mathbf{s}_i)|\mu_Z(\mathbf{s}_i) \sim$ ind. Poisson $(\mu_Z(\mathbf{s}_i))$.
  - $\mu_Z(\mathbf{s}_i) \equiv \lambda(\mathbf{s}_i)E(\mathbf{s}_i)$ is the mean count.
  - $\lambda(\mathbf{s}_i)$ is the **relative risk** for county i, where

$$\lambda(\mathbf{s}_i) = \mu_Z(\mathbf{s}_i)/E(\mathbf{s}_i)$$

- GLM for the mean count:

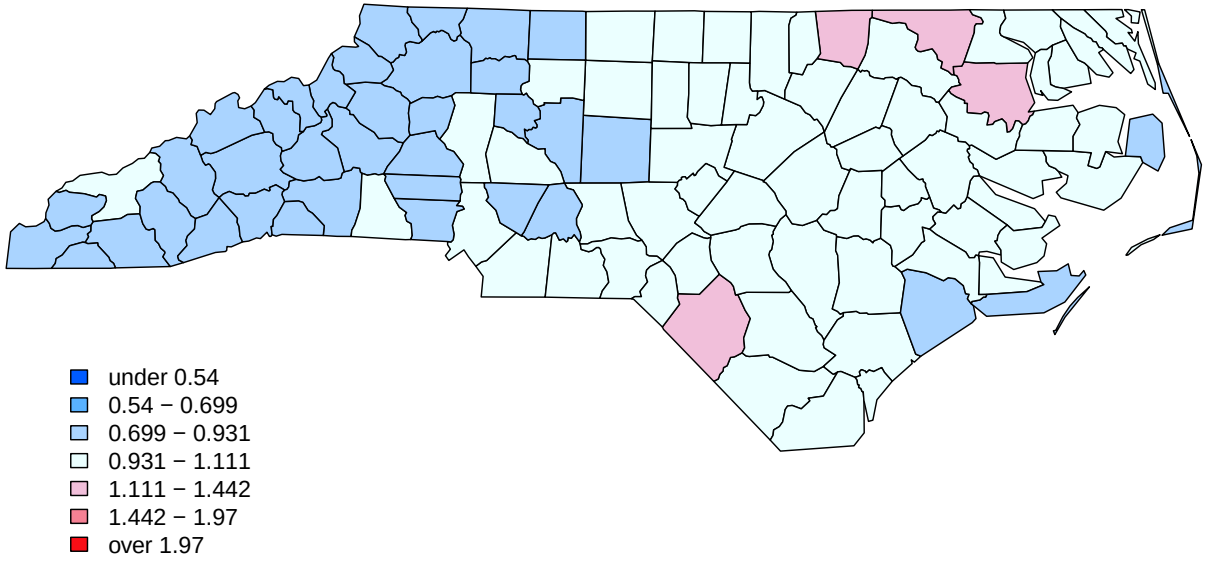
$$\log(\mu_Z(\cdot)) = \log(\lambda(\cdot)E(\cdot)) = \log(E(\cdot)) + \mathbf{x}(\cdot)^\top \boldsymbol{\beta}$$

- The GLM estimate of $\lambda(\cdot)$:

$$\hat{\lambda}_{GLM}(\cdot) = \exp(\mathbf{x}(\cdot)^\top \hat{\boldsymbol{\beta}}_{GLM}).$$

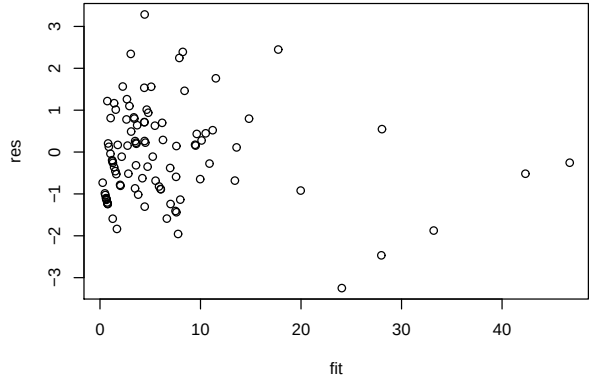
  - $\hat{\boldsymbol{\beta}}_{GLM}$ is obtained by iteratively reweighted least squares.

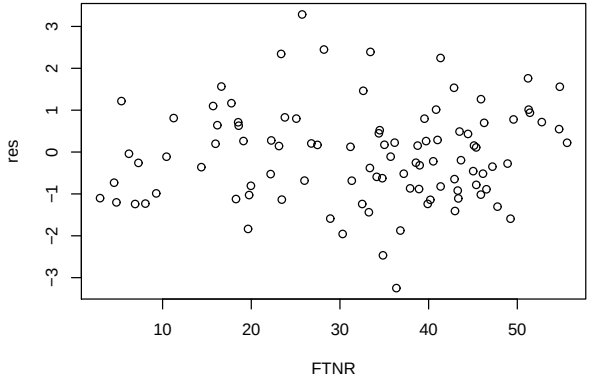
Predicted Relative Risk based on the Standard GLM

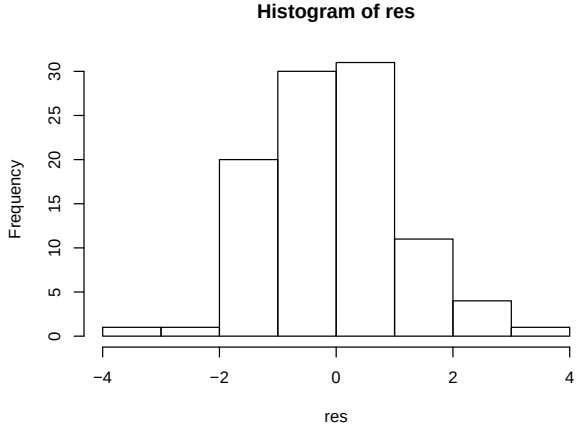


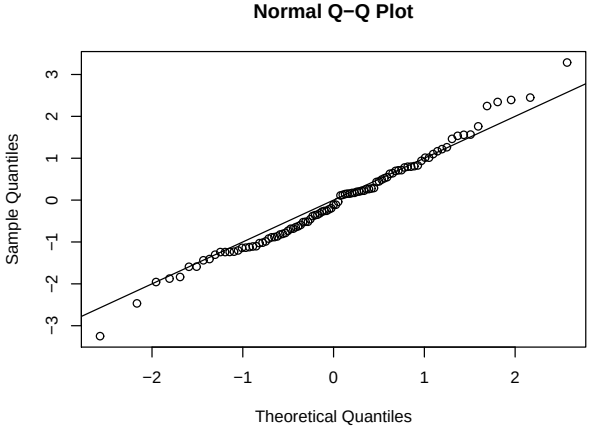
Predicted relative risk, $\hat{\lambda}_{GLM}(\cdot)$, based on the GLM discussed above.

Residual Diagnostic Plots for the Standard GLM Fit

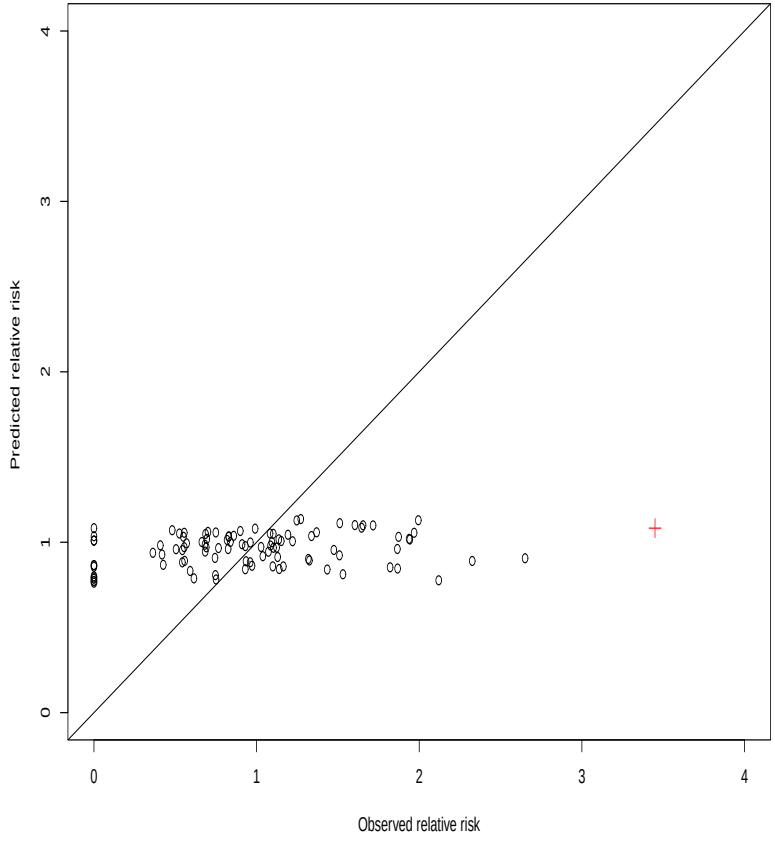








Standard GLM Fit



Predicted relative risk, $\hat{\lambda}_{GLM}(\cdot)$, versus the observed relative risk, $SMR(\cdot)$. County 4 shown as a red "+".

Hierarchical Modeling of Counts Based on the SRE Model

- **Data Model**: $Z(\mathbf{s}_i)|\lambda(s_i) \sim$ ind. **Poisson** $(E(\mathbf{s}_i)\lambda(s_i))$.
  - $\lambda(\mathbf{s}_i)$ is the **relative risk** for county i.
- **Process Model**: $Y(\mathbf{s}_i) = \log(E(\mathbf{s}_i)\lambda(\mathbf{s}_i)) =$
  $\log(E(\mathbf{s}_i)) + \beta_0 + \beta_1 FTNR(\mathbf{s}_i) + E(\mathbf{s}_i)^{-1/2}\left\{\mathbf{S}(\mathbf{s}_i)^\top \boldsymbol{\eta} + \xi(\mathbf{s}_i)\right\}$.
  - $C(\mathbf{s}_i) = \log(E(\mathbf{s}_i))$ is the offset term and $W(\mathbf{s}_i) = E(\mathbf{s}_i)^{-1/2}$ is the weight.
  - $\boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \mathbf{K})$, is 13-dimensional: 12 random effects for the 12 contiguous regions of North Carolina, and one random effect is for county 4, which was previously detected as an outlier (Cressie and Chan [1989]). That is, r=13.
  - For $j = 1, \ldots, 12$, and $i = 1, \ldots, 100$, $S_j(\mathbf{s}_i) = 1$ if the i-th county is in the j-th region, and it is 0 otherwise; and $S_{13}(\mathbf{s}_i) = 1$ if $i = 4$, and it is 0 otherwise.
  - $\xi(\mathbf{s}_i) \sim$ ind. $\mathrm{N}(0, \sigma^2_\xi)$, which models the extra-Poisson variability.
- $\log(\lambda(\cdot)) = \beta_0 + \beta_1 FTNR(\cdot) + E(\cdot)^{-1/2}\left\{\mathbf{S}(\cdot)^\top \boldsymbol{\eta} + \xi(\cdot)\right\}$ **is the log relative risk.**

Statistical Analysis of the Data

- The hidden process $Y(\cdot)$ is spatially dependent, but it is not a MRF. Its spatial dependence comes from a (geostatistical) SRE model; computations are fast because the Sherman-Morrson-Woodbury formula can be used:

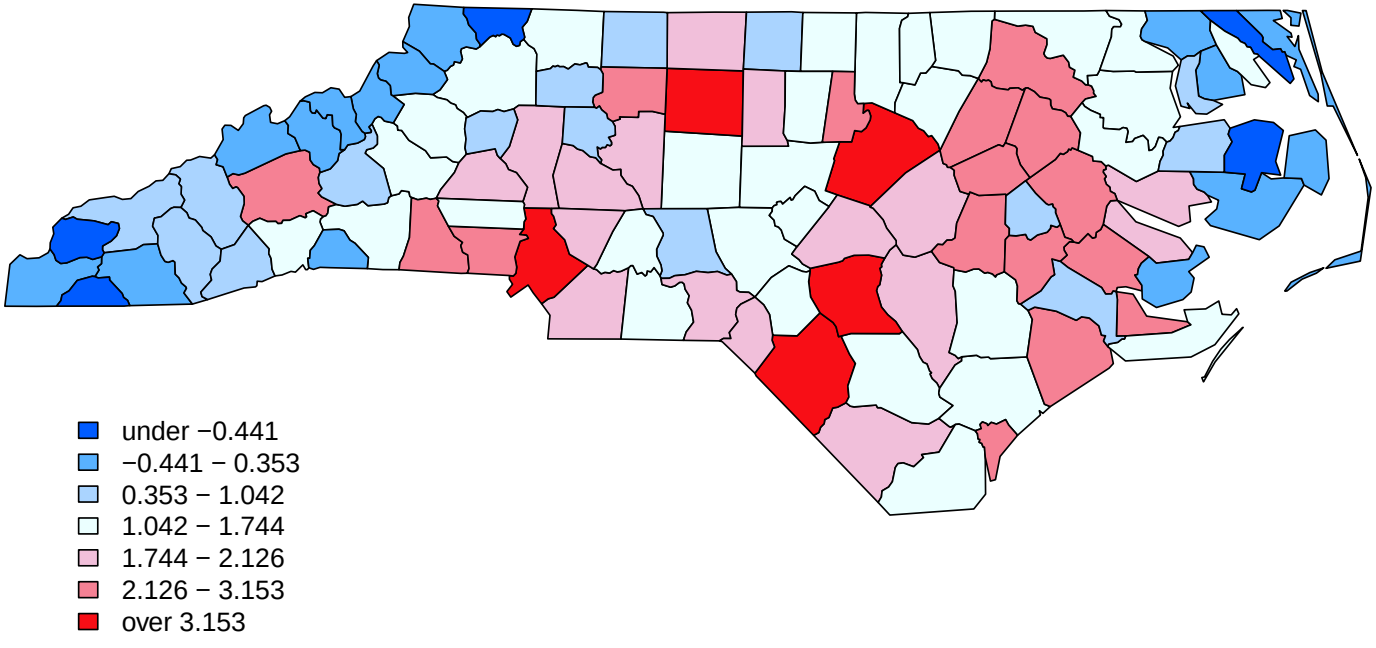
$$(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}.$$

- $\left\{\boldsymbol{\beta}, \mathbf{K}, \sigma^2_\xi\right\}$ **are estimated** using the EM algorithm with a Laplace Approximation in the E-step.
- For inference on $\mathbf{Y}$, we **use the empirical predictive distribution**,

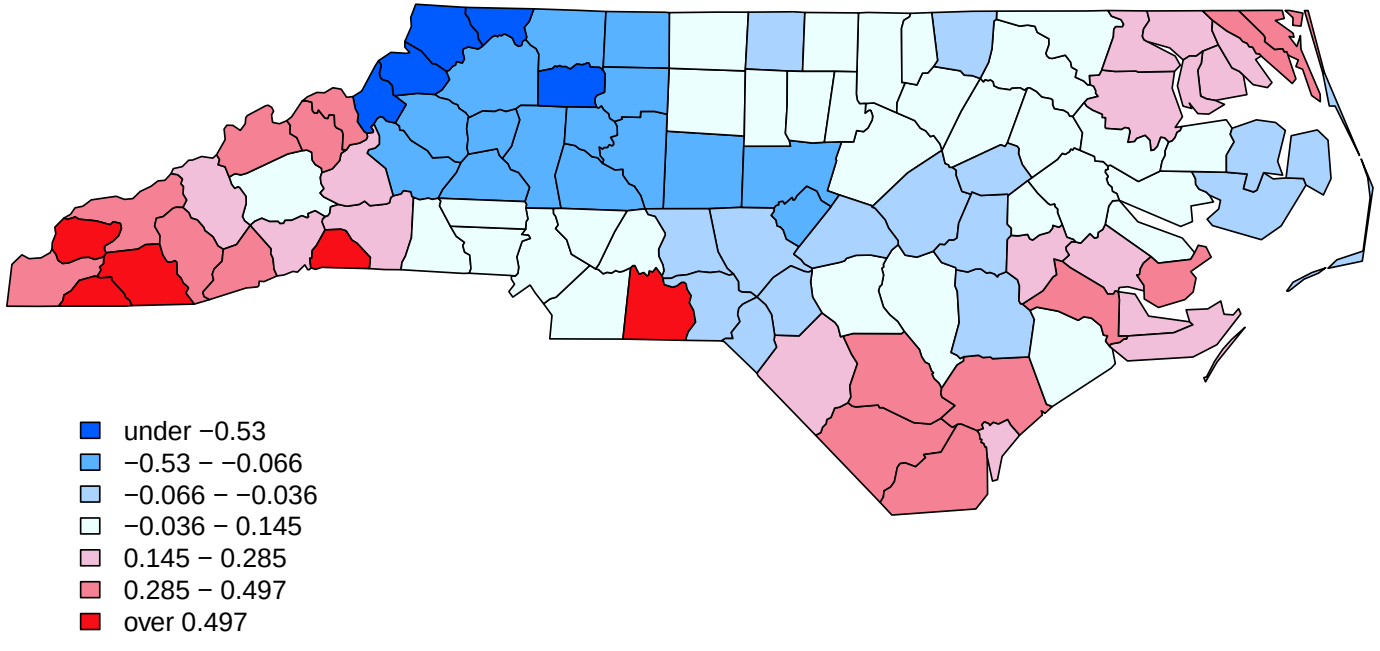
$$[Y(\cdot)|\mathbf{Z}_O, \hat{\boldsymbol{\beta}}_{EM}, \hat{\mathbf{K}}_{EM}, \hat{\sigma}^2_{\xi,EM}],$$

which is obtained using **MCMC**.

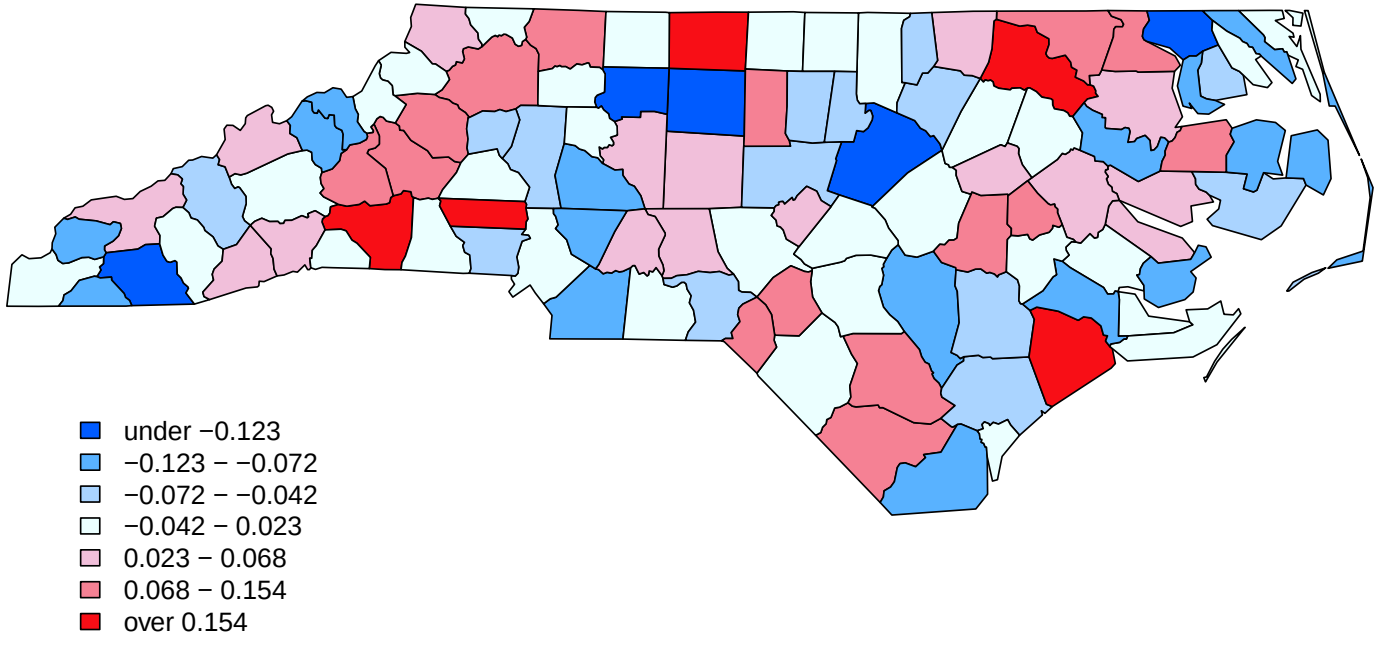
Trend Component : $\log(E(\cdot)) + \mathbf{x}(\cdot)^\top \hat{\boldsymbol{\beta}}_{EM}$



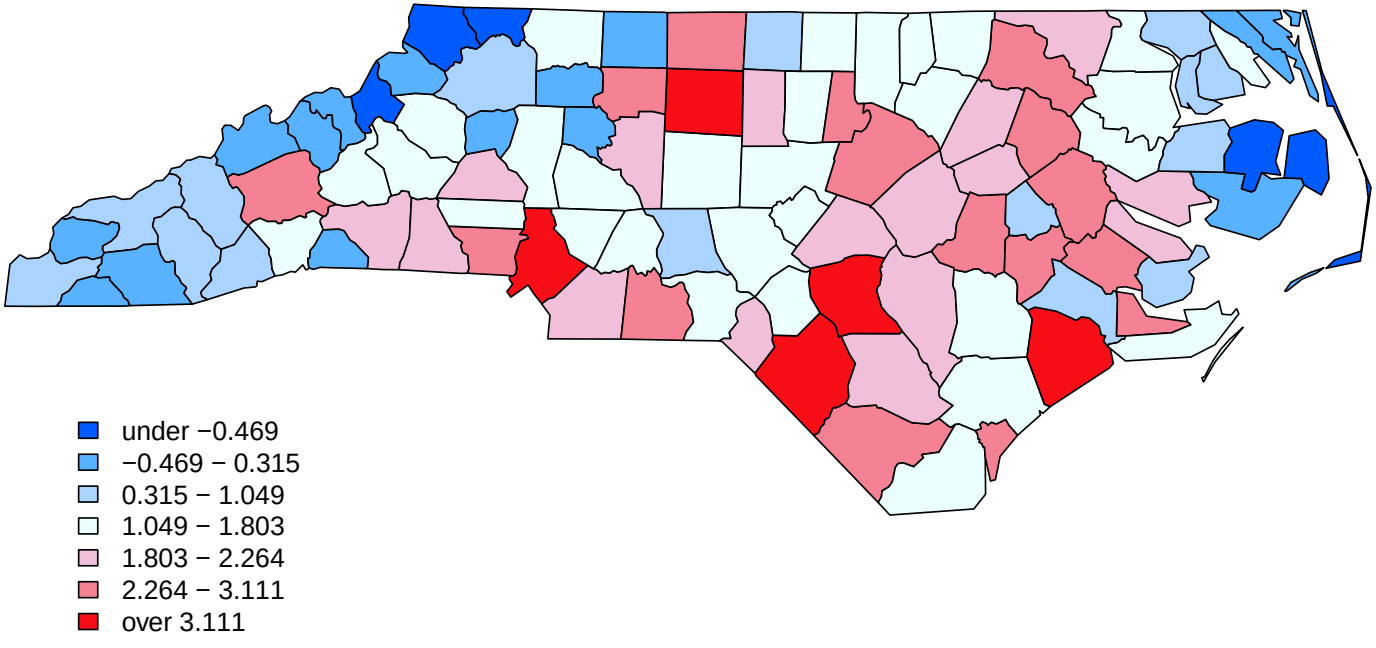
Predictive Mean of the Random Effects : $E(\cdot)^{-1/2}E(\mathbf{S}(\cdot)^{\top}\boldsymbol{\eta}|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}_{EM})$



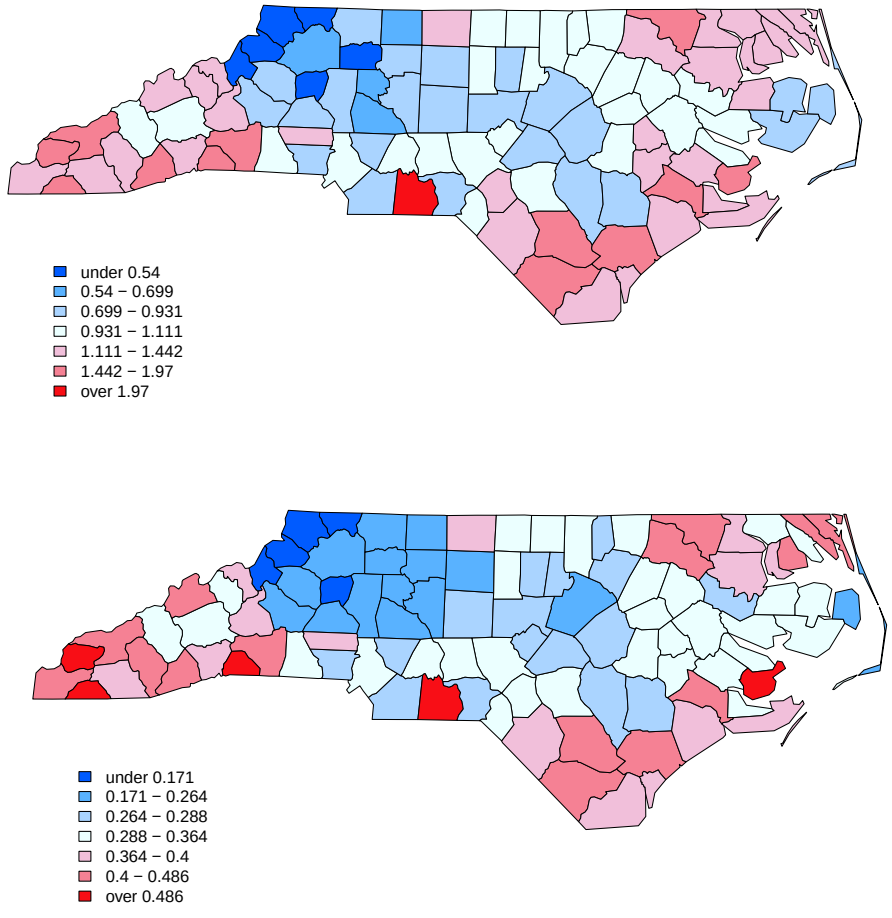
Predictive Mean of Extra Poisson Variability: $E(\cdot)^{-1/2}E(\xi(\cdot)|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}_{EM})$



Predictive Mean of the Sum: $E(Y(\cdot)|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}_{EM})$

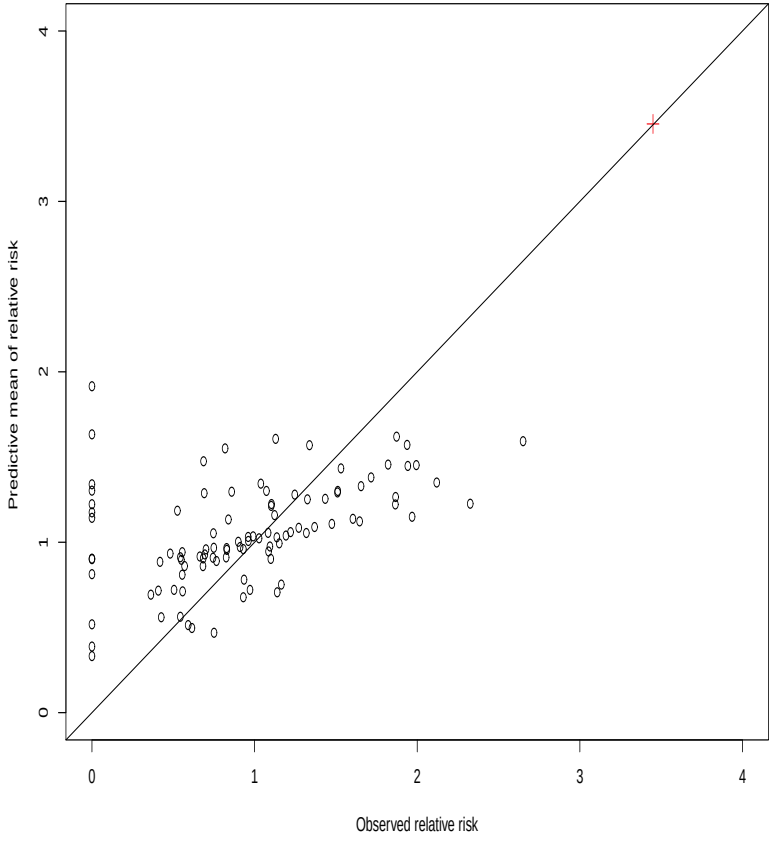


Predictive Mean and S.D. of Relative Risk: $\left[\exp\left(Y(\cdot)\right)/E(\cdot)|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}_{EM}\right]$



Predictive mean (top panel) and predictive standard deviation (bottom panel) of the relative risk.

Model Fit



Predictive mean of the relative risk, $E(\lambda(\cdot)|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}_{EM})$, versus the observed relative risk, $SMR(\cdot)$. County 4 is shown as a red "+". (Counties with zero counts are predicted to be above or below 1, according to their spatial dependence on nearby counties.)

Simulation Study: Computational Efficiency

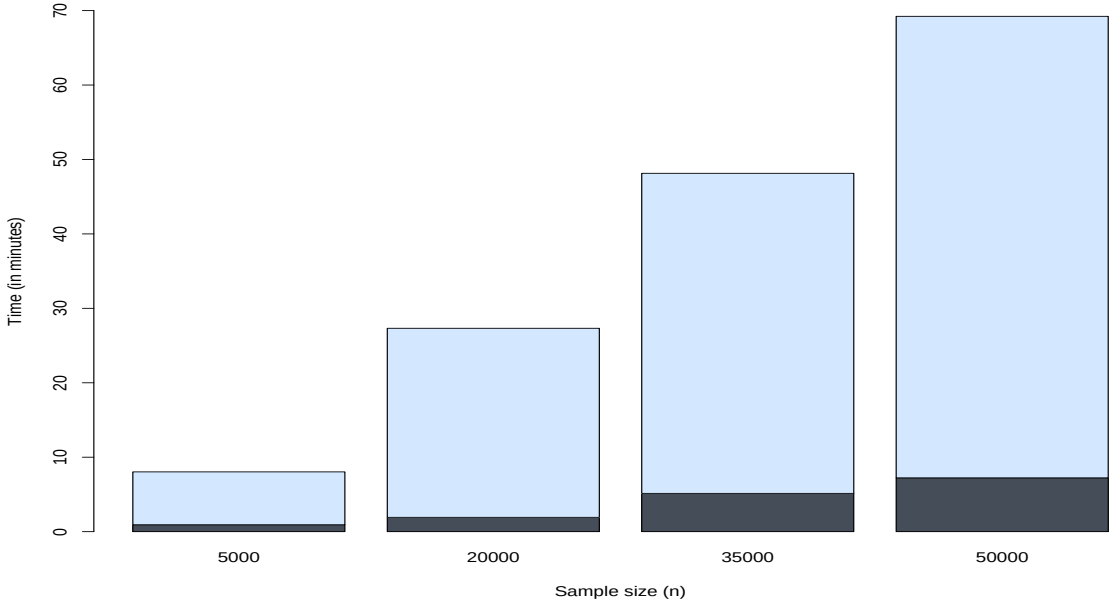
- A simulation experiment was performed for a large dataset ($N = 90,000$, on a 300×300 regular grid).
- The log mean, $\mathbf{Y}$, was simulated from a Gaussian process.
- A realization of counts $\mathbf{Z}$ was simulated with mean $\exp(\mathbf{Y})$, and $n \leq N$ points were sampled as the data ($\mathbf{Z}_O$).

EM Estimation: We estimated the parameters (using EM estimation) from the sampled data $\mathbf{Z}_O$. Computation times were recorded for sample size n.

MCMC: We used an MCMC algorithm to generate $10,000$ samples from the target predictive distribution, after allowing for a burn-in of size $2,000$. Computation times were recorded for sample size n.

The experiment was repeated for four different values of n ($n = 5,000$; $n = 20,000$; $n = 35,000$; and $n = 50,000$).

Computation Time as a Function of n



Plot showing the time taken for EM estimation (dark-gray region) plus the time taken for MCMC implementation (light-gray region), as a function of sample size n. Computation times are clearly $O(n)$.

References

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). *Journal of the Royal Statistical Society, Series B,* **70**:825-848.

Besag, J. E., York, J. C., and Mollié, A. (1991). *Annals of the Institute of Statistical Mathematics,* **43**:1-59.

Cressie, N. (1993). *Statistics for Spatial Data*, revised edn. Wiley, New York, NY.

Cressie, N. and Chan, N. H. (1989). *Journal of the American Statistical Association,* **84**:393-401.

Cressie, N. and Johannesson, G. (2006). In *Australian Academy of Science Elizabeth and Frederick White Conference,* pp 1-11, AAS, Canberra.

Cressie, N. and Johannesson, G. (2008). *Journal of the Royal Statistical Society, Series B,* **70**:209-226.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). *Applied Statistics,* **47**:299-350.

Duncan, W. J. (1944). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Seventh Series,* **35**:660-670.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec., J. P., and Guinet, C. (2006). *Ecological Modeling,* **193**:615-628.

References [cont.]

Lindgren, F., Rue, H., and Lindstrom, J. (2011). *Journal of the Royal Statistical Society, Series B,* **73**:423-498.

Lopes, H. F., Salazar, E., and Gamerman, D. (2008). *Bayesian Analysis,* **3**:759-792.

Lopes, H. F., Gamerman, D., and Salazar, E. (2011). *Computational Statistics and Data Analysis,* **55**:1319-1330.

Robert, C. P. and Casella, G. (2004) *Monte Carlo Statistical Methods,* second edn. Springer, New York, NY.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London, UK: Chapman & Hall/CRC, London, UK.

Sengupta, A., Cressie, N., Frey, R., and Kahn, B. H. (2012). In *Proceedings of the 2012 Joint Statistical Meetings*, Alexandria, VA: American Statistical Association.

Stein, M. L. (2008) *Journal of the Korean Statistical Society,* **37**:3-10

Wikle, C. K. (2002) *Statistical Modelling,* **2**:299-314

Wikle, C. K. (2010). In *Handbook of Spatial Statistics*, pp 107-118, Chapman and Hall/CRC, Boca Raton, FL.

Zhang, H. (2002). *Biometrics,* **58**:129-136.