Space-Time Interaction with Disease Maps

Dongchu Sun

University of Missouri-Columbia

National Institute of Statistical Sciences

and Duke University

Robert K. Tsutakawa

Hoon Kim

Chong Z. He

University of Missouri–Columbia

and
Outline

• Introduction

• Poisson Log-linear Mixed Models

• Default Priors and Propriety of Posterior

• Computation via MCMC

• Model Selection

• Convergence Diagnostic

• Data Analysis via Disease Mapping

• Comments
Example 2

Space-Time Interaction with Disease Maps

Why do we use disease maps?

- To highlight geographic areas with low or high mortality rates of a specific disease, such as lung cancer, and the variability of such rates over a state or country.
- To detect spatial clusterings which may be due to common environmental, demographic or cultural effects shared by neighboring regions.

Figure 1 is a map of crude rates. They are frequencies of annual male lung cancer mortality per 100,000 population by age group, county and period for Missouri.

- The age groups are 45-54, 55-64, 66-74, 75+.
- The periods are 73-77, 78-82, 83-87, 88-92.

There are 115 counties, including St. Louis city, treated as a county distinct.
— The general increase in raw rates over the lower 3 age groups but little increase from the 3rd to 4th age group.
— Some signs of an increase over time for 2 older groups.
— Difficult to detect counties or clusters of counties with high (or low) rates since the extreme rate tend not to appear consistently over county or time.
— The excessive rates typically occur in small counties, where a difference of a few death can have noticeable effect on the crude rates.

The Problems with mapping of crude rates:

• misleading when the size of the population for some of the units are small, resulting in large variabilities in the rates.

• difficult to distinguish chance variability from genuine differences.
The literature on disease mapping grows significantly.

- Breslow & Day (1975)— homogeneous populations of sizes $n_i$, the frequency of deaths $Y \sim \text{Poisson} (n_i p_i)$; factor $p_i$ into several components such as age, sex, and region; then compute MLE of $p_i$.

- Empirical Bayes:
  Manton et al. (1981), Tsutakawa et al. (1985).

- Spatial correlations among neighboring regions:

- Various generalization:
  Tsutakawa (1988),
  Marshall (1991),
  Carlin and Louis (1996),
  Ghosh et al. (1998), etc.
Motivated by the Missouri data, topics of special interest:

a. Demographic effects (age will be used for illustration, though sex and race may also be added.)
b. Spatial effects among different geographic units.
c. Temporal effects on different demographic groups.
d. Regional changes over time.
e. Spatial correlation among regional changes over time.
f. Extra variation not explained by the mixed linear model.
g. Propriety of posterior distribution under noninformative prior distributions.
h. MCMC computing and Bayesian model fitting.

— Most papers on disease mapping focus on $a$ and $b$.

— The extra variation $f$ has been considered by Ghosh et al. (1998) as a term in a generalized linear mixed model.
Poisson Log-linear Mixed Models

Data: $Y_{ijk} \mid p_{ijk} \sim \text{Poisson} \left( n_{ijk} p_{ijk} \right)$,

\[ i = 1, \cdots, I = 115; \ j = 1, \cdots, J = 4 \ (\text{age}) ; \]
\[ k = 1, \cdots, K = 4 \ (\text{time}) . \]

The first stage prior.

— For given \((Z_i, \theta_j, \mu_j, W_{ij}, \delta_0)\),

\[
\log(p_{ijk}) = \theta_j + Z_i + (\mu_j + W_i) (t_k - \bar{t}) + e_{ijk}.
\]

Here, $\theta_j$ and $\mu_j$ are fixed effects;

$Z_i$ and $W_i$ are random effects;

$e_{ijk}$ are independent $N(0, \delta_0)$ errors

(to account for other effects such as race, social or economical status, environmental factor).
The second stage prior.

- The age effects, \( \theta_j \sim N(\xi_{mj}, \delta_{mj}) \) \( j = 1, 2, 3, 4. \)
- Mean slope for age group \( j \), \( \mu_j \sim N(\xi_{sj}, \delta_{sj}). \)
- Choices of the regional effects for \( Z_i (W_i) \):
  - Model 1: simultaneous AR model, (Whittle, 1954)
    
    \[
    Z_i = \rho_1 \sum_{l=1}^{I} C_{il} Z_l + \epsilon_i, \tag{1}
    \]
    
    where \( C_{il} = 1 \) if counties \( i \) and \( l \) are neighbors, 0, else; \( \epsilon_i \ iid \sim N(0, \delta_1) \). Let \( Z = (Z_1, \ldots, Z_I)^t \), adjacency matrix \( C = (C_{il}) \).

    Then (1) is equivalent to \( (I - \rho_1)Z = \epsilon \),
    
    i.e., \( Z \sim \text{MVN} (0, \delta_1 (I - \rho_1 C)^{-2}) \).

  - Model 2: A simultaneous CAR model
    (Clayton & Kaldor, 1987)

    \[
    Z_i|Z_{-i} \sim N(\rho_1 \sum_{l=1}^{I} C_{il} Z_l, \delta_1)
    \]
    
    i.e., \( Z \sim \text{MVN} (0, \delta_1 (I - \rho_1 C)^{-1}) \).
− Model 3: Besag’s CAR model (1991):

\[ f(Z_i|Z_j, j \neq i) = \left( \frac{1}{2\pi \delta_1} \right)^{1/2} \exp \left\{ -\frac{d_i}{2\delta_1} (Z_i - \bar{Z}_i)^2 \right\}, \]

i.e., \( Z \sim \text{MVN}(0, \delta_1(D - C)^{-1}) \), where \( \bar{Z}_i = \sum_j C_{ij} Z_j \), \( d_i = \sum_j C_{ij} \), \( D = \text{diag}(d_1, \ldots, d_I) \).

− Model 3*: Pairwise difference prior (Ghosh, Nataraajan, Stround and Carlin, JASA, 98):

\[ f(Z) \propto \frac{1}{\delta_1^{I/2}} \exp \left[ -\frac{1}{2\delta_1} \sum_{i<j} W_{ij}(Z_i - Z_j)^2 \right], \]

where \( W_{ij} > 0 \) if areas \( i \) and \( j \) are neighbors, 0, else.

− Model 4: a modified CAR model:

\[ f(Z_i|Z_j, j \neq i) = \left( \frac{1}{2\pi \delta_1} \right)^{1/2} \exp \left\{ -\frac{d_i}{2\delta_1} (Z_i - \rho_1 \bar{Z}_i)^2 \right\}, \]

i.e., \( Z \sim \text{MVN} (0, \delta_1(D - \rho_1 C)^{-1}) \).

Will use Model 4 for illustration.

• Prior for \( W \):

\[ W \sim \text{MVN} (0, \delta_2(D - \rho_2 C)^{-1}). \]
The third stage prior.

- \( \delta_l \sim \text{Inverse Gamma} \ (a_l, b_l), l = 0, 1, 2 \)
  \[
  \pi(\delta_l) \propto \delta_l^{a_l-1}e^{-b_l/\delta_l}.
  \]

- \( \rho_1, \rho_2 \ \text{iid.} \ \sim \ \text{uniform} \ (-1, 1) \).
Default Priors and Propriety of Posterior

Need find the 22 hyperparameters
\[(\xi_{mj}, \delta_{mj}), (\xi_{sj}, \delta_{sj}), j = 1, 2, 3, 4; (a_l, b_l), l = 0, 1, 2.\]

- Subjective
  - difficult at the initial stage of analysis
  - different people may have different priors

- Historical data prior
  - Land, Ibrahim and Chen (1997, 1998 a,b,... )
  - He and Sun (1997) used 1994 spring survey data

- Adaptive priors
  - Wasserman et al. (1997)
  - Sun, Tsutakawa, Kim and He (1997)

- Vague or simple invariance priors.

  1. For fixed effects, may use constant priors.
     Let \(\delta_{mj}, \delta_{sj} \rightarrow \infty.\)
2. For random effects, use scale-invariant priors?

\[ \pi(\delta_j) \propto 1/\delta_j, \text{ i.e., } a_j = 0 \text{ and } b_j = 0. \]

**Question:** If we use, is the joint posterior proper?

- Objective, nonsubjective or noninformative prior
  - Sun, Tsutakawa and Speckman (1999, Biomatrika)
  - Sun, Tsutakawa and He (1998)
  - Sun, Speckman and Tsutakawa (1998)

**Question:** If we use, is the joint posterior proper?

**Answer:** No, when \( a_j = 0 \) or \( b_j = 0 \),

\[ p(\delta_j) \propto 1/\delta_j \quad \text{or} \quad p(\log(\delta_j)) \propto 1. \]

Yes, when \( a_j = -0.5 \) and \( b_j = 0 \),

\[ p(\delta_j) \propto 1/\sqrt{\delta_j} \quad \text{or} \quad p(\sqrt{\delta_j}) \propto 1. \]

Yes, when \( a_j = -1 \) and \( b_j = 0 \),

\[ p(\delta_j) \propto 1. \]

— We will begin with this default prior.
Computation via MCMC

For the model we proposed, there are 2083 parameters!

<table>
<thead>
<tr>
<th>$p_{ijk}$</th>
<th>$\theta_j$</th>
<th>$Z_i$</th>
<th>$\mu_j$</th>
<th>$W_i$</th>
<th>$\delta_j$</th>
<th>$\rho_j$</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IJK$</td>
<td>$J$</td>
<td>$I$</td>
<td>$J$</td>
<td>$I$</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1840</td>
<td>4</td>
<td>150</td>
<td>4</td>
<td>150</td>
<td>3</td>
<td>2</td>
<td>2083</td>
</tr>
</tbody>
</table>

- Use Gibbs sampling procedures to find the marginal posterior distributions, posterior moments.

- The full conditional distributions:
  - $\theta_j$, $Z_i$, $\mu_j$, or $W_i$, given others, are normal;
  - $\delta_j$, given others, is inverse gamma;
  - cond. densities of $(\log(p_{ijk})|others)$, and $(\rho_j|others)$ are log-concave (cf. Gilks & Wild (1992) or Berger & Sun (1993)).
Post. Quant. Based on 50,000 Gibbs Cycles

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>$-5.687$</td>
<td>$-5.686$</td>
<td>$0.03892$</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>$-4.508$</td>
<td>$-4.507$</td>
<td>$0.03496$</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>$-3.867$</td>
<td>$-3.865$</td>
<td>$0.03391$</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>$-3.798$</td>
<td>$-3.796$</td>
<td>$0.03503$</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>$-0.00898$</td>
<td>$-0.00875$</td>
<td>$0.008289$</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$0.01210$</td>
<td>$0.01205$</td>
<td>$0.006636$</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>$0.02041$</td>
<td>$0.02035$</td>
<td>$0.005710$</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>$0.03383$</td>
<td>$0.03378$</td>
<td>$0.006087$</td>
</tr>
<tr>
<td>$\delta_z$</td>
<td>$0.01732$</td>
<td>$0.0176$</td>
<td>$0.003201$</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>$0.00386$</td>
<td>$0.00392$</td>
<td>$0.000632$</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>$0.00285$</td>
<td>$0.00288$</td>
<td>$0.000419$</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>$0.00257$</td>
<td>$0.00260$</td>
<td>$0.000364$</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>$0.00272$</td>
<td>$0.00275$</td>
<td>$0.000393$</td>
</tr>
<tr>
<td>$\delta_0$</td>
<td>$0.01305$</td>
<td>$0.01317$</td>
<td>$0.002107$</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>$0.11380$</td>
<td>$0.11200$</td>
<td>$0.02246$</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>$0.01674$</td>
<td>$0.01472$</td>
<td>$0.04129$</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>$0.01365$</td>
<td>$0.01233$</td>
<td>$0.03928$</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>$0.00709$</td>
<td>$0.00602$</td>
<td>$0.03811$</td>
</tr>
<tr>
<td>$\rho_4$</td>
<td>$0.00814$</td>
<td>$0.00706$</td>
<td>$0.03828$</td>
</tr>
</tbody>
</table>
Marginal Posterior Densities

- Figure 2: Posterior densities of $\theta_j$ and $\mu_j$;
  $\theta_j$ tends to be increasing: $\theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4$.
  $\mu_j$ tends to be decreasing.

- Figure 3: Posterior densities of $\delta_0, \delta_1, \delta_2$;

- Figure 4: Posterior densities of $\rho_1$ and $\rho_2$.
  $\rho_1$ tends to be positive, but $\rho_j$ is around 0.
Convergence Diagnostic

In monitoring the convergence of Gibbs sampling, we have run three parallel, MCMC chains with widely different starting values and used graphical monitoring of the chains for a representative subset of the parameters including large, moderate and small mortality rates and other parameters such as $\theta_2, \mu_1, Z_{79}, W_{113}, \delta_2, \rho_2,$ and $\delta_0$.

Figure 7 shows the updated means of Gibbs sampling, with respect to the number of iterations.

It seems that the estimators are quite stable after about 20,000 iterations.
Do Spatial correlations Exist? ($\rho_t = 0$)?

- One conclusion is that spatial correlations among countries exist in terms of $Z_i$, but not in terms of the change over time $W_i$.

- The conclusion is supported by the Bayes factors which were computed using the method of Meng and Wong (1996). For the three reduced models:
  (i) $\rho_1 = 0$, (ii) $\rho_2 = 0$, and (iii) $\rho_1 = \rho_2 = 0$,
  the values of the Bayes factors (in support of the full model) are 31, 912, 2.1, and 62,028, respectively.

- Although there are strong correlations among the $Z_i$, the effect of these correlations on $p_{ijk}$ is relatively small due to the significant effect of $e_{ijk}$.

- Effects of $\rho_1$ and $\rho_2$ on $Z_i$ and $W_i$.

- Effects of $\rho_1$ and $\rho_2$ on $p_{ijk}$. 
Robustness in Terms of Choice of Priors

We compared the hierarchical model by using informative priors and noninformative priors.

The correlation between $\hat{p}_{ijk}$ and $\hat{p}^*_{ijk}$ are 0.999975. The relative error $\frac{|\hat{p}_{ijk} - \hat{p}^*_{ijk}|}{\hat{p}_{ijk}}$ is less than 0.006, while 95.1% of relative errors are less than 0.003. Also, the relative errors of Bayesian estimators of $(Z, W, \theta, \mu, \delta, \rho, \delta_0)$ using noninformative and informative priors are quite small.
Data Analysis via Disease Mapping

- Figure 8: Bayesian estimators (posterior means) of annual male lung cancer mortality rates per 100,000 population by age group, county and period for the State of Missouri.

- Figure 9(a): Bayesian estimators of relative county effects (\( \exp(Z_i) \)). Figure 9(b): Bayesian estimators of mean slope \( W_i \).

- Figure 10: Maps of the extra variations, \( e_{ijk} \), for fixed age group \( j \) and time period \( k \).
Comments

Other models for fitting the data

- We may use Binomial instead of Poisson.
- In general, may use GLM to fit the data.
- Link functions could be $\log[p_{ijk}/(1 - p_{ijk})]$ or other monotonic function.
- Other distributions for $e_{ijk}$ such as log-gamma.

Modeling spatial correlations

- $N(0, \delta(I - \rho_1 C)^{-2})$ (cf. Whittle, 1954).
- $N(0, \delta(I - \rho_1 C)^{-1})$ (cf. Clayton & Kaldor, 1987).
- $N(0, \delta(D - C)^{-})$, Besag et al. (91).
- Pairwise difference prior (Ghosh et al. JASA, 98).
- $N(0, \delta(D - \rho C)^{-1})$, here....
- $C$ may be distance matrix instead of adjacency matrix.
Including other factors

- Account for sex effects.
- Include race effects, grouping, etc.
- Comparing Log-Normal or Gamma extra variations? Besag’s CAR or Simultaneous CAR? (Kim, & and Tsutakawa, 1999a).
- Simultaneous estimation of mortality rates for several diseases will improve accuracy of estimated mortality rates (DeSouza, 1992; Kim, Sun & Tsutakawa 1999b).

Other applications.

- Mortality due to other chronic diseases.
- Other human rates such as those to measure crime, accidents, and high school drop outs.
- Environmental and ecological study.
- Turkey hunting survey (He and Sun, 1997; Woodard, Sun, He and Sheriff, 1998; Woodard and Sun, 1999).
- A homepage on Non-subjective Bayesian Statistical Methodology

  http://www.stat.missouri.edu/~bayes

- To be included in the homepage, send a message to

  dsun@stat.missouri.edu

- For a question on noninformative priors for a specific model, send a message to

  nsbayes@lists.missouri.edu

- To signon the list, email to

  listproc@lists.missouri.edu

  with the following request:

  signon NSBAYES e-address first-name last-name
Noninformative Priors for Linear Mixed Models


Example. Assume that

\[ V_j = \theta_j + e_j, \quad \theta_j \sim \text{iid } N(\mu, \delta_\theta), \quad e_j \sim \text{iid } N(0, \delta_0). \]

Under the scale-invariance prior \( \pi(\mu, \delta_\theta, \delta_0) \propto 1/(\delta_\theta \delta_0) \), the posterior of \( \theta_j \) is improper.

— In principle one may avoid improper posteriors by using diffuse priors, but in practice this may not really solve the problem.

— Because of the increasing computing power, we could apply models of ever greater complexity. This in turn makes it more important to check whether the posterior is proper.
We have consider a general linear mixed model

\[ V = X_1 \theta + X_2 u + e, \quad (2) \]

where \( V \) is an \( n \times 1 \) vector of data, \( X_1 \) and \( X_2 \) are known \( n \times p \) and \( n \times q \) design matrices, \( \text{rank}(X_1) = p \), \( e \) is an \( n \times 1 \) vector of random errors, \( \theta \) is a \( p \times 1 \) vector of fixed effects (parameters), \( u \) is a \( q \times 1 \) vector of random effects (random variables),

\[
\begin{align*}
&u|(\delta_1, \delta_2, \cdots, \delta_r) \sim N_q(\mathbf{0}, A), \\
e|\delta_0 \sim N_n(\mathbf{0}, \delta_0 I_n),
\end{align*}
\]

where \( u = (u'_1, u'_2, \cdots, u'_r)' \), \( u_i \) is \( q_i \times 1 \), \( \sum_{i=1}^{r} q_i = q \), and \( A = \bigoplus_{i=1}^{r} \delta_i B_i^{-1} \),

\[ B_i = (I_{q_i} - \rho_i C_i)^{\xi_i}, \]

where \( \xi_i \geq 0 \) is an integer, \( C_i \) is a known \( q_i \times q_i \) symmetric matrix, whose smallest and largest eigenvalues are \( \lambda_{i1} < 0 \) and \( \lambda_{iq_i} > 0 \), respectively, and \( \rho_i \) is an unknown parameter satisfying \( 1/\lambda_{i1} < \rho_i < 1/\lambda_{iq_i} \).
We assume the following prior densities:

\[
\begin{align*}
    g_\theta(\theta) & \propto 1, \\
    g_i(\delta_i|a_i) & \propto \delta_i^{-(a_i+1)}, \\
    g_0(\delta_0|a_0, b_0) & \propto (\delta_0)^{-(a_0+1)},
\end{align*}
\]

(3)

\[\rho_i \sim \text{cdf } F_i(\cdot) \text{ on } (\lambda_{i1}^{-1}, \lambda_iq_i^{-1}).\]

**Theorem A.** 1. If \( t = q \) or if \( r = 1 \) the following conditions \((a)-(c)\) are necessary and sufficient for the propriety of the posterior dist. of \((\theta, u, \delta_1, \cdots, \delta_r, \rho_1, \cdots, \rho_r, \delta_0)\).

   (a) \( a_i < 0 \),

   (b) \( q_i + 2a_i > q - t \),

   (c) \( n - p + 2a + 2a_0 > 0 \), where \( a = \sum_{i=1}^r a_i \).

2. If \( t < q \) and \( r > 1 \), then Conditions \((a)-(c)\) are sufficient for the propriety of the posterior dist. while necessary conditions result when \((b)\) is replaced with

   (b)’ \( q_i + 2a_i > 0 \).

— Hobert & Casella (96) — when \( \rho_j = 0 \).
Theorem B. Under the assumptions of Theorem A. If $V = (\log(p_{111}), \log(p_{112}), \cdots, \log(p_{IJK}))$, and $Y_{ijk} > 0$, then similar results hold for the joint posterior dist. of $(V, \theta, u, \delta_0, \delta_1, \delta_2, \rho_1, \rho_2)$.

For our mortality data, $n = IJK = 115 \times 4 \times 4 = 1840$, $p = 2J = 8$, $r = 2$, $q = 2I = 230$, and $t = 228$. So $q - t = 2$. 
The Poisson-Normal-CAR model begins with

\[ Y_{ijk} \mid p_{ijk} \sim \text{Poisson}(n_{ijk}p_{ijk}). \]

— The first stage prior for \( p_{ijk}|(Z_i, \theta_j, \mu_j, W_i, \delta_0) \) is

\[ \log(p_{ijk}) = \theta_j + Z_i + (\mu_j + W_i)(t_k - \overline{t}) + e_{ijk}. \]

— The second stage priors are

- \( \pi(\theta_j, \ldots, \theta_J, \mu_1, \ldots, \mu_J) \propto 1; \)
- \( Z| (\delta_1, \rho_1) \sim \text{MVN} (0, \delta_1 A_1^{-2}), \quad A_1 = (D - \rho_1 C); \)
- \( W| (\delta_2, \rho_2) \sim \text{MVN} (0, \delta_2 A_2^{-1}), \)
  where \( W = (W_1, \ldots, W_I)^t \) and \( A_2 = (D - \rho_2 C); \)
- \( \pi(\delta_0) \propto 1/\delta_0. \)

— The third stage priors are

- \( \pi(\delta_1), \pi(\delta_2) \propto 1; \)
- \( \rho_1, \rho_2 \text{ iid } \sim \text{uniform } (-1, 1). \)