Analysis of Marked Point Patterns with Spatial and Non-spatial Covariate Information

Shengde Liang, Bradley P. Carlin, and Alan E. Gelfand

shengdel@biostat.umn.edu, brad@biostat.umn.edu, and alan@stat.duke.edu

Division of Biostatistics, School of Public Health, University of Minnesota

and

Institute of Statistics and Decision Sciences, Duke University
In disease mapping, data are typically presented as aggregated counts over areal regions (counties, zip codes, etc.) ⇒ analyze with areal or lattice models
In disease mapping, data are typically presented as aggregated counts over areal regions (counties, zip codes, etc.) ⇒ analyze with areal or lattice models

Example: Model regional counts as Poisson($E_i e^{\mu_i}$) and assume the $\mu_i$ follow an area-level conditionally autoregressive (CAR) distribution
Background

- In disease mapping, data are typically presented as aggregated counts over areal regions (counties, zip codes, etc.) ⇒ analyze with areal or lattice models

- Example: Model regional counts as Poisson\( \left( E_i e^{\mu_i} \right) \) and assume the \( \mu_i \) follow an area-level conditionally autoregressive (CAR) distribution

- But if precise geographic coordinates are available, the data are properly viewed as a spatial point pattern
In disease mapping, data are typically presented as aggregated counts over areal regions (counties, zip codes, etc.) ⇒ analyze with areal or lattice models

Example: Model regional counts as $\text{Poisson}(E_i e^{\mu_i})$ and assume the $\mu_i$ follow an area-level conditionally autoregressive (CAR) distribution

But if precise geographic coordinates are available, the data are properly viewed as a spatial point pattern

Under a non-homogeneous Poisson process, likelihood for the intensity surface generating the locations is known, but complex...
Background

- In disease mapping, data are typically presented as aggregated counts over areal regions (counties, zip codes, etc.) ⇒ analyze with areal or lattice models

- **Example:** Model regional counts as $\text{Poisson}(E_i e^{\mu_i})$ and assume the $\mu_i$ follow an area-level conditionally autoregressive (CAR) distribution

- **But** if precise geographic coordinates are available, the data are properly viewed as a spatial point pattern

- Under a non-homogeneous Poisson process, likelihood for the intensity surface generating the locations is known, but complex...

- Spatial point process methods and computations both more challenging ⇒ retreat to the Poisson-CAR?...
But we can’t retreat!

- We often have **individual-level covariates** (either “of interest" or “nuisance") we want to incorporate:
  - We seek to compare patterns across certain **treatment** covariates that “mark" the point pattern
  - Other **non-spatial** covariates (e.g., patient characteristics or risk factors) are nuisances
  - Still other **spatial** covariates may be available at either point or areal summary level
But we can’t retreat!

- We often have **individual-level covariates** (either “of interest" or “nuisance") we want to incorporate:
  - We seek to compare patterns across certain **treatment** covariates that “mark" the point pattern
  - Other **non-spatial** covariates (e.g., patient characteristics or risk factors) are nuisances
  - Still other **spatial** covariates may be available at either point or areal summary level

- We propose modeling point patterns **jointly** over geographic and non-spatial nuisance covariate space (precludes aggregation to counts)
But we can’t retreat!

- We often have individual-level covariates (either “of interest" or “nuisance") we want to incorporate:
  - We seek to compare patterns across certain treatment covariates that “mark" the point pattern
  - Other non-spatial covariates (e.g., patient characteristics or risk factors) are nuisances
  - Still other spatial covariates may be available at either point or areal summary level

- We propose modeling point patterns jointly over geographic and non-spatial nuisance covariate space (precludes aggregation to counts)

- Interest lies in certain covariate effects, and the marginal intensity associated with geographic space
But we can’t retreat!

- We often have individual-level covariates (either “of interest" or “nuisance") we want to incorporate:
  - We seek to compare patterns across certain treatment covariates that “mark" the point pattern
  - Other non-spatial covariates (e.g., patient characteristics or risk factors) are nuisances
  - Still other spatial covariates may be available at either point or areal summary level

- We propose modeling point patterns jointly over geographic and non-spatial nuisance covariate space (precludes aggregation to counts)

- Interest lies in certain covariate effects, and the marginal intensity associated with geographic space

- Dependence between treatment-specific intensity surfaces \(\Rightarrow\) multivariate spatial process modeling!
Model-Based Approach

Write the intensity of the process as $\lambda(s)$, where $s$ is a location in a spatial domain $\mathcal{D}$. 
Model-Based Approach

- Write the intensity of the process as $\lambda(s)$, where $s$ is a location in a spatial domain $\mathcal{D}$.

- Cumulative intensity over any block $A$ (say, county or zip code) is $\int_A \lambda(s)\,ds$, which is $\lambda|A|$, with $|A|$ is the area of $A$, if $\lambda(s)$ is free of $s$. 

Analysis of Marked Point Patterns with Spatial and Non-spatial Covariate Information – p. 4/27
Model-Based Approach

- Write the intensity of the process as \( \lambda(s) \), where \( s \) is a location in a spatial domain \( \mathcal{D} \).
- Cumulative intensity over any block \( A \) (say, county or zip code) is \( \int_A \lambda(s) \, ds \), which is \( \lambda |A| \), with \( |A| \) is the area of \( A \), if \( \lambda(s) \) is free of \( s \).
- Likelihood for observed locations \( s_i, \ i = 1, \ldots, n \), is then

\[
L(\lambda(s), s \in D; \{s_i\}_{i=1}^n) = e^{-\int_D \lambda(s) \, ds} \prod_{i=1}^n \lambda(s_i)
\]
Model-Based Approach

- Write the intensity of the process as $\lambda(s)$, where $s$ is a location in a spatial domain $\mathcal{D}$.

- Cumulative intensity over any block $A$ (say, county or zip code) is $\int_A \lambda(s)ds$, which is $\lambda|A|$, with $|A|$ is the area of $A$, if $\lambda(s)$ is free of $s$.

- Likelihood for observed locations $s_i$, $i = 1, \ldots, n$, is then

\[
L(\lambda(s), s \in D; \{s_i\}_{i=1}^n) = e^{-\int_D \lambda(s)ds} \prod_{i=1}^n \lambda(s_i)
\]

- Parametrizing $\lambda(s)$ by $\theta$, adding a prior distribution $p(\theta) \Rightarrow$ posterior distribution $p(\lambda(s; \theta)|\{s_i\})$ for the intensity surface.
Modeling (cont’d)

We typically think of $\lambda(s)$ as a log-Gaussian process (GP) realization, for which the prior might be $p(\lambda(s) \mid \mu(s), \sigma^2, \phi)$, where $\mu(s)$ is the mean, and $\sigma^2$ and $\phi$ are the variance parameters of the GP.
We typically think of $\lambda(s)$ as a log-Gaussian process (GP) realization, for which the prior might be

$$p(\lambda(s) \mid \mu(s), \sigma^2, \phi),$$

where $\mu(s)$ is the mean, and $\sigma^2$ and $\phi$ are the variance parameters of the GP.

We express $\mu(s)$ in part as $z'(s)\beta$, for some spatially referenced covariates $z(s)$.
Modeling (cont’d)

We typically think of $\lambda(s)$ as a log-Gaussian process (GP) realization, for which the prior might be $p(\lambda(s) | \mu(s), \sigma^2, \phi)$, where $\mu(s)$ is the mean, and $\sigma^2$ and $\phi$ are the variance parameters of the GP.

We express $\mu(s)$ in part as $z'(s)\beta$, for some spatially referenced covariates $z(s)$.

Since $\lambda(s)$ is being modeled as a random realization of a spatial process, the integral in the likelihood is stochastic, precluding explicit evaluation. Computational challenges thus include:

- the stochastic integration
- the often large collection of spatial locations (the “big $n$ problem”)
- a prior specification that is only available through finite dimensional distributions.
Wolpert and Ickstadt (1998): fully Bayesian approaches for spatially nonhomogeneous Poisson process data
Really Brief Literature Review

- Wolpert and Ickstadt (1998): fully Bayesian approaches for spatially nonhomogeneous Poisson process data
- Beneš et al. (2002): Bayesian analysis of a log Gaussian Cox process model, assuming $\lambda(s)$ constant over grid cells, and no joint modeling of multiple surfaces or non-spatially indexed covariates
Really Brief Literature Review

- Wolpert and Ickstadt (1998): fully Bayesian approaches for spatially nonhomogeneous Poisson process data

- Beneš et al. (2002): Bayesian analysis of a log Gaussian Cox process model, asssuming $\lambda(s)$ constant over grid cells, and no joint modeling of multiple surfaces or non-spatially indexed covariates

- Diggle (2007): nice review of log Gaussian Cox process modeling both for static and dynamic point patterns
Really Brief Literature Review

- Wolpert and Ickstadt (1998): fully Bayesian approaches for spatially nonhomogeneous Poisson process data

- Beneš et al. (2002): Bayesian analysis of a log Gaussian Cox process model, assuming $\lambda(s)$ constant over grid cells, and no joint modeling of multiple surfaces or non-spatially indexed covariates

- Diggle (2007): nice review of log Gaussian Cox process modeling both for static and dynamic point patterns

- Brix and Diggle (2001) and Duan et al. (2006): spatiotemporal case, using differential equations to drive the evolution of the intensity surface over time.
Really Brief Literature Review

- Wolpert and Ickstadt (1998): fully Bayesian approaches for spatially nonhomogeneous Poisson process data
- Beneš et al. (2002): Bayesian analysis of a log Gaussian Cox process model, assuming \( \lambda(s) \) constant over grid cells, and no joint modeling of multiple surfaces or non-spatially indexed covariates
- Diggle (2007): nice review of log Gaussian Cox process modeling both for static and dynamic point patterns
- Brix and Diggle (2001) and Duan et al. (2006): spatiotemporal case, using differential equations to drive the evolution of the intensity surface over time.

Then there’s the probabilistic discussion in Møller and Waagepetersen (2004), but really not much else out there for fully Bayesian inference...
Jittered residential locations of cases, as well as radiation treatment facilities (RTFs; triangles), northern Minnesota, 1998–2002.

not shown: treatment (BCS vs. mastectomy), age, stage, census variables (education, poverty, race, etc.)
N Minnesota Breast Cancer Data

Two options for surgery:

- **mastectomy**: more invasive and disfiguring, but usually does not require follow-up radiation treatment
- **breast conserving surgery (BCS, or “lumpectomy”)**: requires daily radiation therapy over a 5-week period
N Minnesota Breast Cancer Data

Two options for surgery:

- **mastectomy**: more invasive and disfiguring, but usually does not require follow-up radiation treatment.
- **breast conserving surgery (BCS, or “lumpectomy”)**: requires daily radiation therapy over a 5-week period.

Primary task: compare the BCS and mastectomy point patterns, and account for their dependence when estimating main effects common to both.
N Minnesota Breast Cancer Data

- Two options for surgery:
  - **mastectomy**: more invasive and disfiguring, but usually does not require follow-up radiation treatment
  - **breast conserving surgery (BCS, or “lumpectomy”)**: requires daily radiation therapy over a 5-week period

- **primary task**: compare the BCS and mastectomy point patterns, and account for their dependence when estimating main effects common to both.

- **secondary question**: Are women living in more rural areas (as measured by estimated driving distance to the nearest RTF) more likely to opt for mastectomy?
N Minnesota Breast Cancer Data

Two options for surgery:

- **mastectomy**: more invasive and disfiguring, but usually does not require follow-up radiation treatment
- **breast conserving surgery (BCS, or “lumpectomy”)**: requires daily radiation therapy over a 5-week period

**primary task**: compare the BCS and mastectomy point patterns, and account for their dependence when estimating main effects common to both.

**secondary question**: Are women living in more rural areas (as measured by estimated driving distance to the nearest RTF) more likely to opt for mastectomy?

**Covariates we use**:

- distance to nearest RTF (spatial – exact)
- census tract poverty rate (spatial – areal only)
- patient age and stage (non-spatial)
Modeling with Spatial Covariates

Let $X = \{s_i\}_{i=1}^{n}$ be a set of random locations modeled using a nonhomogeneous Poisson process with intensity $\lambda(s)$.

We take $\lambda(s) = r(s)\pi(s)$, where $r(s)$ is the population density surface at location $s$. In practice, we let $r(s) = \frac{\# \text{ points in } A}{|A|}$ for all $s \in A$. 

Modeling with Spatial Covariates

Let $X = \{s_i\}_{i=1}^n$ be a set of random locations modeled using a nonhomogeneous Poisson process with intensity $\lambda(s)$

- We take $\lambda(s) = r(s)\pi(s)$, where $r(s)$ is the population density surface at location $s$. In practice, we let $r(s) = (\# \text{ points in } A)/|A|$ for all $s \in A$

- Thus $\pi(s)$ is interpreted as a population adjusted (or relative) intensity, which we model on the log scale as

$$
\pi(s) = \exp(\beta' z(s) + w(s)),
$$

where $w(s)$ is a zero-centered stochastic process, and $\beta$ is an unknown vector of regression coefficients.
Modeling with Spatial Covariates

Let \( X = \{ s_i \}_{i=1}^{n} \) be a set of random locations modeled using a nonhomogeneous Poisson process with intensity \( \lambda(s) \)

- We take \( \lambda(s) = r(s) \pi(s) \), where \( r(s) \) is the population density surface at location \( s \). In practice, we let \( r(s) = (\# \text{ points in } A)/|A| \) for all \( s \in A \)

- Thus \( \pi(s) \) is interpreted as a population adjusted (or relative) intensity, which we model on the log scale as

\[
\pi(s) = \exp(\beta' z(s) + w(s))
\]

where \( w(s) \) is a zero-centered stochastic process, and \( \beta \) is an unknown vector of regression coefficients.

- If \( w(s) \) is taken to be a Gaussian process, then the original point process is called a log Gaussian Cox process (LGCP)
Modeling with Spatial Covariates

Likelihood for $\beta$ and $w_D = \{w(s) : s \in D\}$ given $X$: 

$$L(\beta, w_D; X) \propto \exp \left( - \int_D r(s)\pi(s)ds \right) \times \prod_{s_i \in X} r(s_i)\pi(s_i).$$
Modeling with Spatial Covariates

- Likelihood for $\beta$ and $w_D = \{w(s) : s \in D\}$ given $X$:

$$L(\beta, w_D; X) \propto \exp\left(-\int_D r(s)\pi(s)ds\right) \times \prod_{s_i \in X} r(s_i)\pi(s_i).$$

- Adding priors on $w_D$ and $\beta$ leads to

$$p(\beta, w_D|X) \propto L(\beta, w_D; X)p(\beta)p(w_D) \leftarrow \text{intractable!}$$
Modeling with Spatial Covariates

- Likelihood for $\beta$ and $w_D = \{w(s) : s \in D\}$ given $X$:

$$L(\beta, w_D; X) \propto \exp \left( - \int_D r(s) \pi(s) ds \right) \times \prod_{s_i \in X} r(s_i) \pi(s_i).$$

- Adding priors on $w_D$ and $\beta$ leads to

$$p(\beta, w_D | X) \propto L(\beta, w_D; X)p(\beta)p(w_D) \leftarrow \text{intractable!}$$

- In practice, partition $D$ into $A_i$, $i = 1, 2, \ldots, m$, integrate the intensity surface over each $A_i$, and create a Poisson likelihood for the counts given $\lambda(A_i) \leftarrow \text{still requires}$

$$\int_{A_i} r(s) \pi(s) ds,$$

which cannot be done explicitly.
Modeling with Spatial Covariates

Likelihood for $\beta$ and $w_D = \{w(s) : s \in D\}$ given $X$:

$$L(\beta, w_D; X) \propto \exp \left( - \int_D r(s)\pi(s)ds \right) \times \prod_{s_i \in X} r(s_i)\pi(s_i) .$$

Adding priors on $w_D$ and $\beta$ leads to

$$p(\beta, w_D | X) \propto L(\beta, w_D; X)p(\beta)p(w_D) \leftarrow \text{intractable!}$$

In practice, partition $D$ into $A_i$, $i = 1, 2, \ldots, m$, integrate the intensity surface over each $A_i$, and create a Poisson likelihood for the counts given $\lambda(A_i) \leftarrow$ still requires $$\int_{A_i} r(s)\pi(s)ds$$, which cannot be done explicitly.

Could model at the areal level ($\log \pi(A_i)$), but

- precludes use of point level covariate information

$$\pi(A_i) = \int_{A_i} \pi(s) \neq exp(\int_{A_i} (\beta' z(s) + w(s))ds); \text{ latter,}$$

simpler integration could lead to ecological fallacy
Suppose we replace $\int_D \lambda(s)$ with some numerical integration (analytic or Monte Carlo), so we replace $w_D$ with a finite set, say $w^* = \{w(s^*_j), j = 1, 2, \ldots, T\}$. 
Computational Approach

Suppose we replace $\int_D \lambda(s)$ with some numerical integration (analytic or Monte Carlo), so we replace $w_D$ with a finite set, say $w^* = \{w(s_j^*), j = 1, 2, \ldots, T\}$.

Revise the likelihood to

$$L(\beta, w^*, w(s_1), \ldots w(s_n); X)p(w^*, w(s_i), \ldots w(s_n))p(\beta).$$

Now, we only need to work with an $(n + T)$-dimensional random variable to handle the $w$’s, whose prior is just an $(n + T)$-dimensional normal distribution.
Computational Approach

Suppose we replace $\int_D \lambda(s)$ with some numerical integration (analytic or Monte Carlo), so we replace $w_D$ with a finite set, say $w^* = \{w(s_j^*), j = 1, 2, \ldots, T\}$.

Revise the likelihood to

$$L(\beta, w^*, w(s_1), \ldots w(s_n); X)p(w^*, w(s_i), \ldots w(s_n))p(\beta).$$

Now, we only need to work with an $(n + T)$-dimensional random variable to handle the $w$’s, whose prior is just an $(n + T)$-dimensional normal distribution.

We need $z(s)$ at each $s_j^*$, but they are not random so may be interpolated or tiled; we just need to be able to assign a value of $z$ for each $s \in D$. 
Computational Approach

Suppose we replace $\int_D \lambda(s)$ with some numerical integration (analytic or Monte Carlo), so we replace $w_D$ with a finite set, say $w^* = \{w(s_j^*), j = 1, 2, \ldots, T\}$.

Revise the likelihood to

$$L(\beta, w^*, w(s_1), \ldots w(s_n); X)p(w^*, w(s_i), \ldots w(s_n))p(\beta).$$

Now, we only need to work with an $(n + T)$-dimensional random variable to handle the $w$'s, whose prior is just an $(n + T)$-dimensional normal distribution.

We need $z(s)$ at each $s_j^*$, but they are not random so may be interpolated or tiled; we just need to be able to assign a value of $z$ for each $s \in D$.

Our case: distance to RTF = exact, poverty = tiled
Introducing Non-spatial Covariates

Recall two types of non-spatial covariates:
- the “marks” (our case: treatment [MAS vs BCS])
- the “nuisances” (our case: age and cancer stage)

We wish to adjust intensity to reflect age and stage for each treatment group.
Introducing Non-spatial Covariates

Recall two types of non-spatial covariates:
- the “marks” (our case: treatment [MAS vs BCS])
- the “nuisances” (our case: age and cancer stage)

We wish to adjust intensity to reflect age and stage for each treatment group.

Writing the nuisances as continuous, introduce a second argument into the definition of the intensity,

$$\pi(s, v) = \exp \left( \beta' z(s) + w(s, v) \right).$$

This is a surface over the product space $D \times \mathcal{V}$ (i.e., geographic space by covariate space)
Introducing Non-spatial Covariates

Recall two types of non-spatial covariates:
- the “marks” (our case: treatment [MAS vs BCS])
- the “nuisances” (our case: age and cancer stage)

We wish to adjust intensity to reflect age and stage for each treatment group.

Writing the nuisances as continuous, introduce a second argument into the definition of the intensity,

$$\pi(s, v) = \exp \left( \beta' z(s) + w(s, v) \right).$$

This is a surface over the product space $D \times V$ (i.e., geographic space by covariate space)

Interest is in the “marginal” spatial intensity associated with $\pi(s, v)$, based on data $\{(s_i, v_i), \ i = 1, 2, \ldots, n\}$. 
Introducing Non-spatial Covariates

In the interest of separating $s$ and $v$ in the model, for now let us write $w(s, v) = w(s) + u(v)$

– model $w(s)$ as above (GP), but take $u(v) = v'\alpha$
Introducing Non-spatial Covariates

- In the interest of separating $s$ and $v$ in the model, for now let us write $w(s, v) = w(s) + u(v)$
  - model $w(s)$ as above (GP), but take $u(v) = v'\alpha$

- Separability allows $\pi(s, v)$ to factor as $\exp\{u(v)\}\pi(s) \Rightarrow$
  - can study the spatial intensity scaled by the nuisances
Introducing Non-spatial Covariates

- In the interest of separating $s$ and $v$ in the model, for now let us write $w(s, v) = w(s) + u(v)$
  - model $w(s)$ as above (GP), but take $u(v) = v' \alpha$

- Separability allows $\pi(s, v)$ to factor as $\exp\{u(v)\} \pi(s)$ \Rightarrow can study the spatial intensity scaled by the nuisances

- Separable additive marked log relative intensity:

  \[
  \log \pi_k(s, v) = \beta_{0k} + z'(s) \beta_k + v' \alpha_k + w_k(s) .
  \]

- $\beta_{0k}$ capture the global mark effects

- spatially-referenced covariates and nuisances can differentially affect the mark-specific intensities

- $w_k$ provide a different GP realization for each $k$. 
Introducing Non-spatial Covariates

In the interest of separating $s$ and $v$ in the model, for now let us write $w(s, v) = w(s) + u(v)$

– model $w(s)$ as above (GP), but take $u(v) = v'\alpha$

Separability allows $\pi(s, v)$ to factor as $\exp\{u(v)\}\pi(s)$ ⇒ can study the spatial intensity scaled by the nuisances

Separable additive marked log relative intensity:

$$\log \pi_k(s, v) = \beta_{0k} + z'(s)\beta_k + v'\alpha_k + w_k(s).$$

- $\beta_{0k}$ capture the global mark effects
- spatially-referenced covariates and nuisances can differentially affect the mark-specific intensities
- $w_k$ provide a different GP realization for each $k$.

Sensible reduced models: $w_k(s) = w(s)$, $\beta_k = \beta$, $\alpha_k = \alpha$. 

Introducing Non-spatial Covariates

- In the interest of separating \( s \) and \( v \) in the model, for now let us write \( w(s, v) = w(s) + u(v) \)
  - model \( w(s) \) as above (GP), but take \( u(v) = v'\alpha \)

- Separability allows \( \pi(s, v) \) to factor as \( \exp\{u(v)\}\pi(s) \Rightarrow \) can study the spatial intensity scaled by the nuisances

- Separable additive marked log relative intensity:

\[
\log \pi_k(s, v) = \beta_0 k + z'(s)\beta_k + v'\alpha_k + w_k(s).
\]

- \( \beta_0k \) capture the global mark effects
- spatially-referenced covariates and nuisances can differentially affect the mark-specific intensities
- \( w_k \) provide a different GP realization for each \( k \).

- Sensible reduced models: \( w_k(s) = w(s), \beta_k = \beta, \alpha_k = \alpha. \)

- Or: Add multiplicative interaction between \( z(s) \) and \( v \)
Introducing Non-spatial Covariates

The intensity associated with our separable additive model is

$$\lambda_k(s, v) = \exp(\beta_0 + v' \alpha_k) \times r(s) \exp(z'(s) \beta_k + w_k(s)) .$$

⇒ latter part is the natural marginal spatial intensity!
Introducing Non-spatial Covariates

The intensity associated with our separable additive model is

\[ \lambda_k(s, v) = \exp(\beta_{0k} + v'\alpha_k) \times r(s) \exp(z'(s)\beta_k + w_k(s)) . \]

\[ \Rightarrow \text{latter part is the natural marginal spatial intensity!} \]

Possible dependence among the \( w_k(s) \) surfaces \( \Rightarrow \) multivariate Gaussian process over the \( w_k \).
Introducing Non-spatial Covariates

The intensity associated with our separable additive model is

\[ \lambda_k(s, v) = \exp(\beta_{0k} + v' \alpha_k) \times r(s) \exp(z'(s) \beta_k + w_k(s)) . \]

⇒ latter part is the natural marginal spatial intensity!

Possible dependence among the \( w_k(s) \) surfaces ⇒ multivariate Gaussian process over the \( w_k \).

Cross-covariances \( \Gamma_w(s, s') = [Cov(w_1(s), w_2(s'))] \) must be specified carefully so that process realizations remain positive definite:

- separable forms (easy!)
Full Likelihood Specification

Letting \( \{(s_{ki}, v_{ki}), i = 1, 2, ... n_k\} \) be the locations and nuisances associated with the \( n_k \) points having mark \( k \), the likelihood becomes

\[
\prod_k \exp \left( - \int_D \int_V \lambda_k(s, v) dv ds \right) \times \prod_k \prod_{s_{ki}, v_{ki}} \lambda_k(s_{ki}, v_{ki}) .
\]
Full Likelihood Specification

Letting \( \{(s_{ki}, v_{ki}), i = 1, 2, \ldots n_k\} \) be the locations and nuisances associated with the \( n_k \) points having mark \( k \), the likelihood becomes

\[
\prod_k \exp \left( - \int_D \int_V \lambda_k(s, v) dv ds \right) \times \prod_k \prod_{s_{ki}, v_{ki}} \lambda_k(s_{ki}, v_{ki}).
\]

Under our separable additive model, this becomes

\[
\prod_k \exp \left( -q(\beta_{0k}, \alpha_k) \int_D r(s) \exp(z'(s)\beta_k + w_k(s)) ds \right) \times \prod_k \prod_{i=1}^{n_k} \left( \exp(\beta_{0k} + v_{ki}\alpha_k) r(s_{ki}) \exp(z'(s_{ki})\beta_k + w_k(s_{ki})) \right),
\]

where \( q(\beta_{0k}, \alpha_k) = \int_V \exp(\beta_{0k} + v'\alpha_k) dv \).

Additive form in \( z(s) \) and \( v \) results in single integrals.

Our approximations permit use of the same set of integration grid points \( s_{j}^* \) for each \( k \).
Computational Issues

Our \((n + T)\)-dimensional normal distributions above require working with inverses and determinants of very large covariance matrices – the “big \(N\)” problem in geostatistics!
Computational Issues

Our \((n + T)\)-dimensional normal distributions above require working with inverses and determinants of very large covariance matrices – the “big \(N\)” problem in geostatistics!

Our (simple) approach: replace \(w(s)\) by an approximation \(\tilde{w}(s)\) in a lower-dimensional subspace.
Computational Issues

Our \((n + T)\)-dimensional normal distributions above require working with inverses and determinants of very large covariance matrices – the “big \(N\)” problem in geostatistics!

Our (simple) approach: replace \(w(s)\) by an approximation \(\tilde{w}(s)\) in a lower-dimensional subspace.

Specifically, at any point \(s_0\), replace the original process \(w(s_0)\) in the likelihood by the predictive process \(\tilde{w}(s_0) = E(w(s_0)|w^*)\), where

\[
w^* = [w_k(s_j^*)]_{k,j} \sim MVN(0, \Gamma^*(\theta))
\]

is a realization of \(w(s)\) over an arbitrary set of knots \(S^* = \{s_1^*, \ldots, s_m^*\}\) (Banerjee et al., 2007).
Computational Issues

Our \((n + T)\)-dimensional normal distributions above require working with inverses and determinants of very large covariance matrices – the “big \(N\)” problem in geostatistics!

**Our (simple) approach:** replace \(w(s)\) by an approximation \(\tilde{w}(s)\) in a lower-dimensional subspace.

Specifically, at any point \(s_0\), replace the original process \(w(s_0)\) in the likelihood by the predictive process \(\tilde{w}(s_0) = E(w(s_0)|w^*)\), where

\[ w^* = [w_k(s_j^*)]_{k,j} \sim MVN(0, \Gamma^*(\theta)) \]

is a realization of \(w(s)\) over an arbitrary set of knots \(S^* = \{s_1^*, \ldots, s_m^*\}\) (Banerjee et al., 2007).

Fit the model using a Gibbs sampler to update the parameters in \(\pi(s, v)\) as well as the random effects at the knots.
Application to N Minn Data

Location-specific covariates:
- \( z_1(s) \), the log standardized distance to nearest RTF
- \( z_2(s) \), the poverty rate in the census tract containing \( s \)

Non-location-specific covariates:
- \( v_1 \), the patient’s stage at diagnosis (1 if “late" [regional or distant], and 0 otherwise)
- \( v_2 \), the patient’s age at diagnosis
Application to N Minn Data

Location-specific covariates:
- \( z_1(s) \), the log standardized distance to nearest RTF
- \( z_2(s) \), the poverty rate in the census tract containing \( s \)

Non-location-specific covariates:
- \( v_1 \), the patient’s stage at diagnosis (1 if “late" [regional or distant], and 0 otherwise)
- \( v_2 \), the patient’s age at diagnosis

Assume population density \( r(s) \) is constant within tract (2000 census data)
Application to N Minn Data

Location-specific covariates:
- $z_1(s)$, the log standardized distance to nearest RTF
- $z_2(s)$, the poverty rate in the census tract containing $s$

Non-location-specific covariates:
- $v_1$, the patient’s stage at diagnosis (1 if “late" [regional or distant], and 0 otherwise)
- $v_2$, the patient’s age at diagnosis

Assume population density $r(s)$ is constant within tract (2000 census data)

Though we have exact spatial coordinates, the figures on next few slides are presented at tract level, since image-contour plots require preliminary interpolation (say, via interp or MBA in R) which can be misleading over our irregular areal grid.
Non-spatial covariates & crude intensity

For mastectomy (top row) and BCS (bottom row),

- **left:** observed median age
- **middle:** observed proportion of late diagnosis;
- **right:** observed log-relative intensity (count divided by population).
Population and spatial covariates

- **left:** population density by tract
- **middle:** log-standardized distance to nearest RTF by tract
- **right:** poverty rate by tract

Note the Red Lake Indian Reservation in the poverty map
Model fitting

- Priors:
  - \( \sigma \sim U(0, 100), \tau \sim U(-0.999, 0.999) \)
  - \( \phi \sim U(0.5, 25) \Rightarrow \) correlation between the two closest sites, \( \exp(-\phi d_{min}) \), varies from 0.01 to 0.91
Model fitting

- **Priors:**
  - $\sigma \sim U(0, 100), \ \tau \sim U(-0.999, 0.999)$
  - $\phi \sim U(0.5, 25) \Rightarrow$ correlation between the two closest sites, $\exp(-\phi d_{\text{min}})$, varies from 0.01 to 0.91

- **Results:** Full bivariate spatial model outperforms univariate spatial and nonspatial (GLMM, Bivariate GLMM) models in terms of DIC
Model fitting

- Priors:
  - $\sigma \sim U(0, 100)$, $\tau \sim U(-0.999, 0.999)$
  - $\phi \sim U(0.5, 25) \Rightarrow$ correlation between the two closest sites, $\exp(-\phi d_{min})$, varies from 0.01 to 0.91

- Results: Full bivariate spatial model outperforms univariate spatial and nonspatial (GLMM, Bivariate GLMM) models in terms of DIC.

- Table (next slide) gives fixed effect estimates, where second rows give differential effect in the BCS group.

- Age does not affect the relative intensity of mastectomies, but older women are somewhat less likely to choose BCS.

- The random effects models identify distance to the nearest RTF as significant for mastectomy, but the simple GLM estimate misses this.
## Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>model</th>
<th>biv spat res</th>
<th>BGLMM</th>
<th>GLM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS</td>
<td>int</td>
<td>-8.11(0.070)</td>
<td>-8.12 (0.072)</td>
<td>-8.09(0.049)</td>
</tr>
<tr>
<td></td>
<td>log-dist</td>
<td>-0.15(0.040)</td>
<td>-0.15 (0.047)</td>
<td>-0.02(0.023)</td>
</tr>
<tr>
<td></td>
<td>poverty</td>
<td>-1.92(0.565)</td>
<td>-1.85 (0.576)</td>
<td>-1.16(0.366)</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>0.01(0.013)</td>
<td>0.01 (0.012)</td>
<td>0.02(0.012)</td>
</tr>
<tr>
<td></td>
<td>late</td>
<td>-0.64(0.046)</td>
<td>-0.64 (0.047)</td>
<td>-0.65(0.046)</td>
</tr>
<tr>
<td>BCS</td>
<td>int</td>
<td>-0.02(0.086)</td>
<td>-0.03 (0.085)</td>
<td>0.00(0.070)</td>
</tr>
<tr>
<td></td>
<td>log-dist</td>
<td>-0.14(0.040)</td>
<td>-0.14 (0.043)</td>
<td>-0.14(0.034)</td>
</tr>
<tr>
<td></td>
<td>poverty</td>
<td>-0.91(0.673)</td>
<td>-0.83 (0.667)</td>
<td>-1.06(0.551)</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>-0.04(0.020)</td>
<td>-0.04 (0.020)</td>
<td>-0.04(0.019)</td>
</tr>
<tr>
<td></td>
<td>late</td>
<td>-1.02(0.083)</td>
<td>-1.02 (0.083)</td>
<td>-1.01(0.085)</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.83(0.053)</td>
<td>0.82 (0.058)</td>
<td>–</td>
</tr>
</tbody>
</table>
Fitted intensity surfaces, full model

For mastectomy (top row) and BCS (bottom row), and assuming the mean age and an early diagnosis,

- **left**: log-relative intensity without spatial residuals
- **middle**: spatial residuals
- **right**: complete log-relative intensity surfaces
Fitted log-relative intensity surfaces

- **left column:** the two spatial covariates alone encourage higher (darker) values in the south and near the population centers

- **middle column:** compensating residual activity in several rural and suburban tracts

- **right column:** resembles spatially smoothed versions of the corresponding “raw” maps (though direct comparison is not really possible due to age and stage adjustment)

Similarity of the mastectomy and BCS spatial residual maps \( \Leftrightarrow \) fairly large estimated \( \hat{\tau} \) (nonspatial correlation)
Discussion

Summary: Extended the LGCP model to accommodate
- covariates that are spatially referenced
- individual-level covariates that mark the process
- individual-level “nuisance” risk factors

Fitted areal-level log-relative intensity maps now adjusted for the non-spatially varying covariates
**Discussion**

**Summary:** Extended the LGCP model to accommodate
- covariates that are spatially referenced
- individual-level covariates that mark the process
- individual-level “nuisance” risk factors

Fitted areal-level log-relative intensity maps now adjusted for the non-spatially varying covariates

**Future work:**
- Extending to the case of three-way interactions, e.g., mark by space by individual covariates
- Imprecision in the (typically rural) addresses?
- **Space-time** point pattern analysis: separable versus non-separable models for log-intensity, etc.
- **Wombling** to determine statistically significant boundaries in the residual or fitted relative intensity surface (Liang, Banerjee and Carlin, 2008)......
Wombling for Spatial Point Processes

- **Areal**: assuming that spatial residuals are regional, we set \( w_k(s) = w_{ki} \), if \( s \in \text{region } i \), and assign them a multivariate conditionally autoregressive (MCAR) distribution. Following Lu and Carlin (2005), define the boundary for mark \( k \) as those segments having large values of \( E(\Delta_{ij,k}|\text{Data}) \), where \( \Delta_{ij,k} = |w_{ki} - w_{kj}| \).
Wombling for Spatial Point Processes

**Areal:** assuming that spatial residuals are regional, we set \( w_k(s) = w_{ki} \), if \( s \in \text{region } i \), and assign them a multivariate conditionally autoregressive (MCAR) distribution. Following Lu and Carlin (2005), define the boundary for mark \( k \) as those segments having large values of \( E(\Delta_{ij,k} | Data) \), where \( \Delta_{ij,k} = |w_{ki} - w_{kj}| \).

**Point-level:** For a mean-squared differentiable surface \( Y(s) \) and any open curve \( C \) in the domain, Banerjee and Gelfand (2006) define the wombling measure of \( C \) as

\[
\int_C D_{n(s)} Y(s) dv = \int_C \langle \nabla Y(s), n(s) \rangle dv ,
\]

i.e. the total gradient along \( C \), where \( \langle \cdot, \cdot \rangle \) is the inner product, \( n(s) \) is the normal direction to \( C \), and \( D_{n(s)} Y(s) \) is the directional derivative along \( n(s) \).
Boundaries (top 20% of $E(\Delta_{ij,k}|Data)$) shown as dark edges for mastectomy (left) and BCS (right)

- **Cook County** (far NE corner) has statistically higher log intensity than Lake County (adjacent to the W)
- **Red Lake Reservation** (resembles a “P” rotated 90 degrees clockwise) significantly separated from all but one neighboring tract
Residual wombling, point-level case

top: Estimated residual surfaces (left MAS, right BCS)
bottom: Mean predicted gradient surfaces

White lines are candidate boundaries; all are “Bayesianly significant" at the 0.05 level.

THE END (at last!)