2011

Bayesian Point Event Modeling in Spatial and Environmental Epidemiology: A Review

Andrew Lawson
Medical University of South Carolina

MUSC Division of Biostatistics and Epidemiology Working Papers

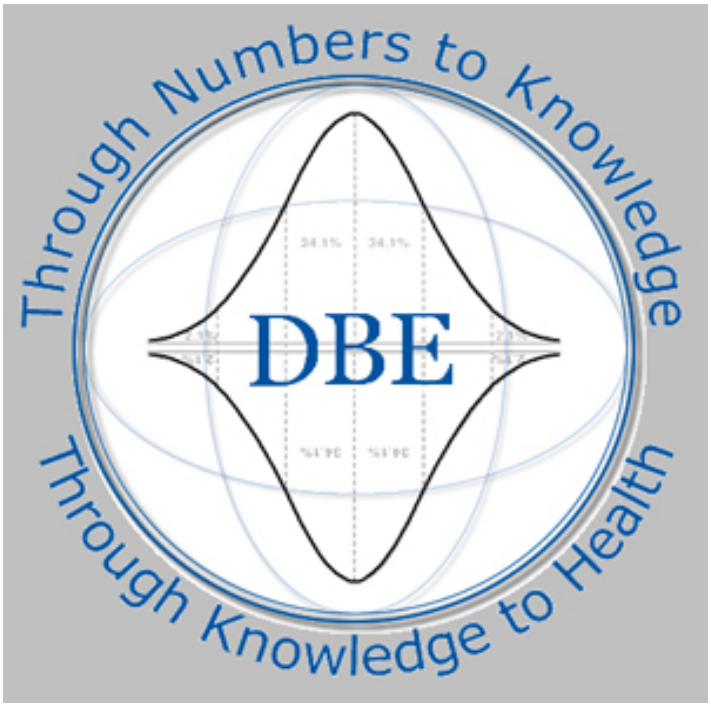
Paper Number/ Resource Identifier: 11-004

Date: July 18, 2011

MUSC Author(s): Andrew B. Lawson

Paper Title: Bayesian Point Event Modeling in Spatial and Environmental Epidemiology: a review

Complete Author List: Andrew B. Lawson



MUSC
MEDICAL UNIVERSITY of SOUTH CAROLINA

Bayesian Point Event Modeling in Spatial and Environmental Epidemiology: a review

Andrew B. Lawson

Division of Biostatistics & Epidemiology

College of Medicine

Medical University of South Carolina

Charleston

SC

USA

Abstract

This paper reviews the current state of point event modeling in spatial epidemiology from a Bayesian perspective. Point event (or case event) data arise when geo-coded addresses of disease events are available. Often this level of spatial resolution would not be accessible due to medical confidentiality constraints. However, for the examination of small spatial scales it is important to be capable of examining point process data directly. Models for such data are usually formulated based on point process theory. In addition, special conditioning arguments can lead to simpler Bernoulli likelihoods and logistic spatial models. Goodness-of-fit diagnostics and Bayesian residuals are also considered. Applications within putative health hazard risk assessment, cluster detection, and linkage to environmental risk fields (misalignment) are considered.

Key words

Case event; point event; point process; logistic; Bayesian; Poisson process; Cox process; log Gaussian process; semi-parametric; putative hazard; predictor fields

1. Introduction

Bayesian modeling of small area count data has seen considerable advances in recent times (see e.g. Besag et al, 1991, Knorr-held, 2000, and Lawson, 2009 amongst others). This is partly due to the availability of such data and its ready application. However, there has been little development of equivalent modeling of geo-referenced address data. Address data naturally form point processes in space and the modeling of such data is often either prohibited due to confidentiality restrictions on health records or because the models are less familiar to the practitioners. A recent technical review of these point process models and approaches is found in Møller and Waagpeteresen (2007). It is the purpose of this review to examine the application of Bayesian methods where geo-referenced health outcome data is the natural format. Most of this paper will focus on spatial models and spatial data, but in a later section (Section 7) I will briefly examine the issues related to spatio-temporal case event data and modeling.

The rest of the paper is laid out as follows. In the next sections I discuss the sources of data, the nature of geo-referenced health data and the use of control diseases to represent population effects. In the subsequent section I examine basic models for this data and define the Bayesian context for such models. In particular I examine the heterogeneous Poisson Process (HPP) model and its extension to a log Gaussian Cox Process (LGCP) and spatial cluster processes. Following that I review specific application areas where Bayesian model have or can be applied. Finally I summarize the current state of development and suggest areas for future work. An appendix with program code is also provided for a selection of the examples.

2. Data Sources

Geo-coded address data, also known as *case event* data, is found where a point location is used to represent the residential address of a case of disease. Hence a sample of street addresses of disease cases yields case event data. Whereas if only a postal code or zip code or other arbitrary area is known for the cases then (usually) only a count of disease can be analyzed within that area. The distinction between case event and count data is blurred by the fact that counts are just a spatially-aggregated

form of case event data. However, when point locations are available a substantially different form of analysis must be performed.

Usually geo-coded address data are found where studies focus on relatively small spatial windows. When considering, for example, the effects of an environmental pollution source on surrounding communities, it may be important to consider a scale of analysis that is defined within a study window of only a few kilometers or miles. Because the scale of postal or census regions is often large by comparison it would not be feasible in these cases to examine the aggregated counts of disease due to the inherent smoothing of the risk variation. In short the risk variation would be smoothed out by aggregation.

The use of such case event data has many advantages and some disadvantages. First of all the data is held at the individual case level and so could be augmented by individual covariates or predictors thought to be relevant in the analysis. On the other hand the use of a residential address may be questioned in that it is used to represent the assumed point at which exposure has taken place. How appropriate this location is will depend on prior understanding of the exposure process and disease etiology.

Case event data can be obtained from local departments of health but are usually subject to stringent confidentiality restrictions. Usually individual level data requests from health departments are subject to concerns about identification of patients. Hence address data must be handled with care. A map at a fine level of resolution could lead to identification of individual patients (even when individual identifiers are not held in the data set). Hence the mapping of such data and its public presentation may also be subject to restrictions in addition to de-identification of patients before analysis.

3. Control Diseases

As in the analysis of count data in aggregate spatial units, it is important to account for the population underlying the case event data. All disease events occur within a population which has a spatially varying density and also varies from place to place in its demographic composition. Because some population groups have greater or lesser susceptibility to disease than others it could be important to account for this variation so that apparent concentration of disease are not mistaken for population peaks. In essence, larger/denser populations give rise to greater disease incidence and this effect is a fundamental feature of a *linear* model of disease risk. Figure 1 displays the spatial distribution of case events (larynx cancer) and controls (respiratory cancer) for a well known example: larynx cancer incidence in Lancashire NW England 1973-1984 (Diggle,1990).

4. Models

Some basic notation that will be useful in the rest of this paper follows. Define a study window as a bounded area within which we observe events. The window is denoted by *T* and its area by *A* . A realization of cases of disease is found within the window *T.* Unlike a random sample, a realization is a complete enumeration of all events (cases) within an area (in this case *T*). Assume that there are *m* observed cases. This set of events has locations defined by $S:\{s_1,.....,s_m\}$. Here, *s* represents the two dimensional geographic coordinate of the case. This is usually a residential address as we usually associate exposure with residence, unless an occupational etiology were to be the focus. In veterinary studies, the set *S* might be siting locations for diseased animals, for example.

We also assume that we observe within *T* a set of *n* control locations: $C:\{c_1,.....,c_n\}$. This set consists of a geographical control for the case outcomes. This could be another disease or the non-diseased portion of the population at risk. For example we could observe maternal residences of babies born with abnormalities and as control we could examine all normal birth locations for the same time period and area.

It is often useful to derive a composite binary variable from the concatenation of case and control realizations. Define the $N=m+n$ set of locations where the first m are cases and the next n are controls as: $S_c:\{S\cup C\}=\{s_{c1},.....,s_{cN}\}$. Define a binary vector $y_i,\ i=1,...,N$ where the first m are cases and the last *n* are controls thus

$$y_i=\begin{cases}1 & if\ \ \mathrm{i}\in\{1,...,m\}\\ 0 & otherwise\end{cases}.$$

This vector is now a binary outcome variable, associated with the S_c locations, and can be considered as the outcome of choice and modeled within a logistic formulation (as will be discussed in detail in the next section).

5. Bayesian approach

Bayesian modeling has as a fundamental ingredient the specification of a likelihood for the observed data. In addition it requires that parameters in models are regarded as stochastic and are assigned prior distributions. The product of the likelihood and prior distributions for parameters yields a posterior distribution for the parameters after suitable normalization. Inference about models is largely based on this posterior distribution. By modelling both the observed data and any unknown parameter or other unobserved effects as random variables, the hierarchical Bayesian approach to statistical analysis provides a cohesive framework for combining complex data models and external knowledge or expert opinion (Lawson and Banerjee, 2009). In general, the data model, which is used in the likelihood, can be denoted $p(y|\boldsymbol{\theta})$ for observation *y*

and a set of parameters $\boldsymbol{\theta}$*,* and then the likelihood for *N* observations is given by $L(\mathbf{y}\,|\,\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i\,|\boldsymbol{\theta})$, for a sample of $\mathbf{y}' = \{y_1, \ldots\ldots, y_N\}$ of size *N* . The assumption is made that given the parameters then the data are conditionally independent.

Modeling event locations: *the Heterogeneous Poisson Process Likelihood*

In our application, the random quantity is the address locations of cases. Hence we need to define our likelihood as $L(\mathbf{s}\,|\,\boldsymbol{\theta}) = \prod_{i=1}^{m} p(s_i\,|\boldsymbol{\theta})$ for a suitably defined $p(s\,|\,\boldsymbol{\theta})$. Once a choice is made for $p(s\,|\,\boldsymbol{\theta})$ then suitable prior distributions must be chosen for the parameters $\boldsymbol{\theta}$. It is conventional in the analysis of case events to assume that the events form a realization from a heterogeneous Poisson process (HePP) . This process describes the distribution of points within a spatial region and the process is governed by an intensity function which is defined for any location *s* as $\lambda(s\,|\,\boldsymbol{\theta})$. This function is strictly positive and yields a local 'rate' of events anywhere within the study area. It is this function that can be modeled via a likelihood. Formal properties of these processes are described in Diggle (2003), ch 5. Two forms of likelihood are associated with this process: conditional and unconditional. Here I only consider the likelihood conditional on *m* events being found in a study area. In this case the probability of an event at a specific location *s* is just

$p(s\,|\,\boldsymbol{\theta}) = \dfrac{\lambda(s\,|\,\boldsymbol{\theta})}{\Lambda_T(\boldsymbol{\theta})}$ where $\Lambda_T(\boldsymbol{\theta}) = \int_T \lambda(u\,|\,\boldsymbol{\theta})du$. Note that $\Lambda_T(\boldsymbol{\theta})$ is just a normalizing constant and it is the integral over the study area of the intensity function $\lambda(s\,|\,\boldsymbol{\theta})$. Hence in this situation we have the log likelihood:

$$l(\mathbf{s}\,|\,\boldsymbol{\theta}) = \sum_{i=1}^{m} log(\lambda(s_i\,|\,\boldsymbol{\theta})) - m\,log\,\Lambda_T(\boldsymbol{\theta}) \qquad (\mathrm{m1})\,.$$

This likelihood can be used for inference within a Bayesian modeling framework, assuming that $\lambda(s\,|\,\boldsymbol{\theta})$ is suitably specified. Usually when we consider spatial epidemiological examples we need to consider the population at risk in relation to the case density and it is common for the intensity function to have 2 parts: a model part and a background part.

For example, we can define $\lambda(s\,|\,\boldsymbol{\theta}) = \lambda_0(s).\lambda_1(s\,|\,\boldsymbol{\theta})$ where $\lambda_0(s)$ represents background population variation and $\lambda_1(s\,|\,\boldsymbol{\theta})$ is the model component. If we assume that there are no parameters to estimate within $\lambda_0(s)$ (ie that it is fixed or given) then in that case we have

$$l(\mathbf{s}\,|\,\boldsymbol{\theta}) = \sum_{i=1}^{m} log(\lambda_0(s_i)) + \sum_{i=1}^{m} log(\lambda_1(s_i\,|\,\boldsymbol{\theta})) - m log\,\Lambda_T(\boldsymbol{\theta})$$

$$= \sum_{i=1}^{m} log(\lambda_1(s_i\,|\,\boldsymbol{\theta})) - m log\,\Lambda_T(\boldsymbol{\theta}) \qquad \text{(m2).}$$

These likelihoods (m1, or m2) are not based on standard distributions and to use them for inference one would have to evaluate the spatial integral $\Lambda_T(\boldsymbol{\theta})$. This of course is not difficult as it is 2 dimensional, and it can be evaluated by using Monte Carlo integration, or even numerical integration schemes if the area were regular such as rectangular, for example. Note that attempts have been made to allow the fitting of these likelihoods in standard software packages, via the use of a Poisson distribution, which has a likelihood similar to that of the HePP form (Berman and Turner,1992; Lawson, 1992). By choosing a numerical integration scheme the likelihood m1 can be written

$$\sum_{i=1}^{m} w_i [\, I_i\, log(\lambda(s_i\,|\,\boldsymbol{\theta})) - \lambda(s_i\,|\,\boldsymbol{\theta})] \qquad \text{(m3)}$$

$$\text{as } \Lambda_T(\boldsymbol{\theta}) = \int_T \lambda(u\,|\,\boldsymbol{\theta})du \approx \sum_{i=1}^{m} w_i \lambda(s_i\,|\,\boldsymbol{\theta})$$

and the likelihood m3 is a weighted Poisson likelihood for indicator variable I_i with mean $\lambda(s_i\,|\,\boldsymbol{\theta})$ and with integration weights $\{w_i\}$ and an indicator vector where $I_i = 1/w_i$. The weights could be Dirichlet tile areas around the case event locations, for example. The choice of weights is really crucial for the accuracy of the integral approximation. Usually, the accuracy of this approximation is not good unless extra dummy weights are added to the sum. Note that m3 can be programmed using the zeroes or ones trick in WInBUGS as an arbitrary likelihood and so a Bayesian hierarchical model can be based on m3 (see Appendix 1). Of course outside of WinBUGS it is possible to simply evaluate m2 with a suitable integration scheme where $\Lambda_T(\boldsymbol{\theta}) \approx \sum_{i=1}^{m} w_i \lambda(s_i\,|\,\boldsymbol{\theta})$. Once a likelihood is specified within WInBUGS it is then possible to extend our model to include covariates (via linear or spline functions), random effects, both spatially correlated and uncorrelated simply by suitable specification of $\lambda(s_i\,|\,\boldsymbol{\theta})$ or $log(\lambda(s_i\,|\,\boldsymbol{\theta}))$.

The above discussion assumes that $\lambda_0(s)$ is known or given of course. An early solution to the problem when $\lambda_0(s)$ is unknown was to use a nonparametric intensity estimator of the control data to yield $\hat{\lambda}_0(s)$ and to use that as a '*plug in'* within m2, with $\lambda(s_i\,|\,\boldsymbol{\theta}) = \hat{\lambda}_0(s_i)\lambda_1(s_i\,|\,\boldsymbol{\theta})$ and therefore a profile likelihood would result. The resulting

inference was found to be sensitive to the smoothing employed to estimate $\hat{\lambda}_0(s_i)$ (Lawson and Williams, 1994). Another approach to this problem is to change the outcome data that is used and to convert the problem into a simpler binary regression. This is discussed in a later section.

5.1 Intensity specification

A fundamental part of the modeling process is the specification of the intensity to be modeled. In spatial epidemiology we are mainly interested in focusing on the specification of $\lambda_1(s_i|\boldsymbol{\theta})$ or $log(\lambda_1(s_i|\boldsymbol{\theta}))$. How this is achieved will in part depend on the application. However one can in general consider the generalized mixed model formulation where

$$log(\lambda_1(s_i|\boldsymbol{\theta})) = x'_{1i}\boldsymbol{\beta}_1 + f(x'_{2i}) + z'_i\boldsymbol{\gamma} \qquad \text{I1}.$$

Here the first term is a linear predictor for covariates (x'_1), the second term consists of potentially non-linear functions (splines) of covariates (x'_2) and finally the third term is a collection of random effects (z') with inclusion vector $\boldsymbol{\gamma}$. The covariates in the first term could include the spatial coordinates of the *i* th location or functions such as distance or direction from a known location (as is used in putative health hazard detection). In addition personal covariates could be included such as age, or gender of the case. The choice of which covariates to include in the non-linear term depends on application. It may be that a 2D field of measurements are to be related to the events and this must be smoothed over space. In that case, a 2D spline might be considered. An example would be pollution measurements at spatial sites that must be interpolated to the point locations of events. Finally we may consider that additional random variation in the case outcomes occurs. This may possibly be due to unobserved confounding. The confounders could be unknown or unmeasured. First an individual level frailty term could be considered whereby individual differences in disease response is allowed. Second, we could consider spatial random heterogeneity which could modulate the intensity. This could be uncorrelated or correlated (spatially structured). We use the acronyms UH and CH to distinguish these effects. Hence within a hierarchical modeling framework we would have a hierarchy of effects specified via prior distributions just as in conventional Bayesian models.

A more concrete example would be useful. In a study of sources of air pollution risk on respiratory asthma cases, we are concerned with the distance effect from a putative pollution source (incinerator), we are also concerned to make allowance for the age of the cases, as well as for the spatial distribution of a competing risk (PM2.5 measured at a network of sites). We want to allow for confounding also. We assume that confounding could be spatially correlated.

A possible formulation would be

$$log(\lambda_1(s_i \mid \boldsymbol{\theta})) = \beta_0 + G_i + \beta_3 age_i + f(pm_i) + u_i \qquad I2$$
$$where\ G_i = \beta_1 log[1 + exp(\beta_2 d_i)]$$

Here, the link to the source is via distance (d_i) in an additive-multiplicative form, with covariate age_i and 2d spline on pm_i (interpolated to the case sites), and finally u_i which is a spatially correlated random effect. This example of course would require the addition of prior distributions for all parameters and effects ($\beta_0, \beta_1, \beta_2, \beta_3,$ parameters in $f(.), u_i$) to complete the Bayesian specification. We consider the full Bayesian specification in a later section.

5.2 Modeling the binary outcome vector

In the above we assumed that $\lambda_0(s)$ is known or can be estimated to provide a profile likelihood. However this is often not available or could lead to smoothing problems.

An alternative formation of the point process model is to consider instead, the joint realization of cases and controls and to model the binary label on the joint vector. Hence instead of directly modeling the locations, we model the binary labels ($y_i,\ i = 1,...,N$) on the locations. It is known form Point process theory that the joint distribution of cases and controls is a HePP also with intensity $\lambda_0(s) + \lambda_0(s)\lambda_1(s \mid \theta)$, assuming that the controls are HePP with intensity $\lambda_0(s)$. From this definition, then we can define the conditional probability of a case ($y_i = 1$) at a given location as

$$Pr(y_i = 1) = \frac{\lambda_0(s)\lambda_1(s \mid \theta)}{\lambda_0(s) + \lambda_0(s)\lambda_1(s \mid \theta)} = \frac{\lambda_1(s \mid \theta)}{1 + \lambda_1(s \mid \theta)} \qquad (Pr1).$$

This is just a logistic probability for a binary outcome variable. In the process of derivation we have removed the nuisance term ($\lambda_0(s)$) and so don't have to estimate it now, and we have derived a conventional probability model (Bernoulli). Niote that spatial dependience can appear in the model but it is addressing the labeling rather than the locations themselves. Diggle and Rowlingson (1994) first proposed this approach and its extensions into semi-parametric modeling (Kelsall and Diggle, 1998). Under a Bernoulli data model the binary outcome variable can be treated as any other binary variable and so a hierarchical model can be constructed which in this case can be termed a Bayesian spatial logistic model. Hence we have

$$y_i \sim Bern(p_i)$$
$$\text{logit}(p_i) = \lambda(s_i \mid x'_{1i}, x'_{2i}, z'_i, \theta).$$

Hence a logistic linear spatial model would arise if $\lambda(s_i \mid x'_{1i}, x'_{2i}, z'_i, \theta)$ is defined by a linear predictor in covariates and random effects. Instead a logistic spline spatial model arises when $\lambda(s_i \mid x'_{1i}, x'_{2i}, z'_i, \theta) = f(x'_{2i})$. Appendix 2 provides an example of WinBUGS code for a logistic linear spatial model with

$$\lambda(s_i \mid x'_{1i}, x'_{2i}, z'_i, \theta) = (1 + exp\{-\gamma_1 d_i\}).exp\{\gamma_0 + \gamma_2 x_{1i} + v_i + w_i\}$$

where

d_i is distance to a putative point source

x_{1i} is age of the individual

v_i is an uncorrelated (UH) random effect

w_i is a spatially correlated (CH) random effect

The focus in the analysis is to make inference about the distance effect (γ_1) making allowance for individual age (x_{1i}) and unobserved confounders ($v_i + w_i$). The choice of random effect and their prior distributions here could be important as we have individual locations where the labels are observed. An uncorrelated effect at the individual level maybe straightforward to model but correlated (CH) effects may not be. Correlation between fixed locations in this case must be considered. There are two basic approaches that could be adopted for this CH specification: 1) a full MVN covariance model where the spatial covariance between locations is a function of their distance separation, and 2) conditional autoregressive (CAR) model whereby a neighborhood relations are assumed between locations. In the first case, the proper prior distribution is defined as

$$w \sim N_m(\mu, \Sigma)$$

where the covariance is $\Sigma_{ij} = \tau_w^{-1} f(d_{ij})$, and d_{ij} is the distance between the i th and j th locations and $\mu = \underline{0}$.

Usually, Σ has parameters that must be estimated in the model. For example, a common simple model is the exponential covariance where $\Sigma_{ij} = \tau_w^{-1} exp(-\alpha d_{ij})$ so that the variability is modeled by τ_w^{-1} and the distance dependence is modeled by α.

The second approach attempts to avoid the computational burden of estimation of covariance parameters and instead considers a CAR alternative. CAR models require a simple neighborhood specification for their spatial structure and if such can be found for a set of locations then this could be a possible model. There are a number of ways one could define neighboring points to a given point: an arbitrary distance threshold could be used for example. An alternative known as ‘natural neighbor hoods’ (Sibson, 1981;

Preparata and Shamos, 1988) uses a tiling of the points and assumes that any adjacencies contain neighboring points. Often a Dirichlet tessellation is specified as this has the property that all locations within a tile are closer to the tile point than to any other point. Hence the tessellation defines 'territories' for the point set. Figure 2 displays the tessellation of a set of arbitrary points. This was created by the R function `DELDIR` which is designed to compute such tessellations and triangulations. One disadvantage of the tessellation is that it does not have support at the edges of the study region. Instead it is always possible to compute the dual of the tessellation, the Delauney triangulation. This is also shown on Figure 2 and is formed by joining the points with a common boundary. In this example the Delauney neighbors (points joined by a triangle) are given by the coordinates X: 0.1 0.2 0.4 0.45 0.6 0.3; and Y: 0.1 0.3 0.7 0.2 0.7 0.4, with the number of neighbors defined as 2, 4, 3, 4, 3, 4. In this way it is possible to define an adjancy matrix and hence a CAR random effect could be assumed. Note that no knowledge of the exact location of the points is required, only the labels of the adjacent points. Appendix 2 displays the WinBUGS code for this model for the larynx cancer dataset (1036 case-control locations) and the dataset is available as detailed in the Appendix.

5.3 Prior Distributional Choices

So far I have only considered a small range of data types and approaches to likelihood definition. Once a likelihood is assumed, whether it is (m1) - (m3) or based on (Pr1), with the intensity specified, then it is important to specify prior distributions for all parameters and random effects. It is one of the great advantages of a Bayesian modeling approach that it is possible to place model structure at different levels of model hierarchy and in this way can avoid complicated likelihood specification. In fact in spatial statistics a major concern is the introduction of spatial correlation into a model. By allowing correlation to be placed in prior distributions a Bayesian model can exploit conditional independence down the hierarchy to allow independent likelihood contributions.

Assuming we have the general form $log(\lambda_1(s_i \mid \boldsymbol{\theta})) = x_{1i}'\boldsymbol{\beta}_1 + f(x_{2i}') + z_i'\boldsymbol{\gamma}$, then the different parameters and effects can be treated as follows. First, we usually try to choose non-informative prior distributions if possible. For regression parameters ($\boldsymbol{\beta}_1$) within linear predictors it is commonly assumed that they can each be independently assigned a zero mean Gaussian distribution with precision τ_* i.e. $\beta \sim N(0, \tau_\beta^{-1})$. The precision is just the inverse of the variance and it measures how concentrated the distribution is. It is conventional to use the precision (rather than the variance) for Gaussian distributions within Bayesian analyses. For uncorrelated random effects a similar specification is made with zero mean and specified precision: $v_i \sim N(0, \tau_v^{-1})$. It is conventional to

consider the combined specification of parameter prior distribution and precision hyper-prior distribution. It is important to specify a prior distribution for a precision parameter (especially for random effects) as

$\tau_* \sim Ga(a,b)$ with $E(\tau_*) = a / b$
so that
$\tau_*^{-1} \sim IG(a,b)$,

where IG(.) stands for inverse gamma distribution. Often a reasonably non-informative Gamma with large variance is chosen for the precision with a=0.001,b=0.001 or even an exponential specification such as Ga(1,0.026) (Wakefield, 2007). Other specifications can be considered as well, such as 'close-to' exponential: a=2.0, b=0.5. Unfortunately while gamma prior distributions are a natural choice for precisions they can sometimes lead to singularities at zero and also can be difficult to sample within McMC samplers. As an alternative, Gelman (2006) suggested using half Cauchy prior distributions and sd-uniform variants. For example, the specification $\tau_*^{-1/2} \sim U(0,c)$ has been recommended and used extensively in WinBUGS help examples. The choice of c is usually in the range c= 5 to c=10. However, even a value of 5 leads to a very liberal range for the variance parameter and so small values of c are sometimes utilized to ensure convergence, especially in models with random effects.

5.4 Cox Process formulation

A Cox process results when the intensity specified is regarded as stochastic. This lends itself to a hierarchical formulation as within a Bayesian formalism the conditioning occurs within a hierarchy and at lower levels of the hierarchy conditional independence results. Hence, in the Cox process we have an intensity that has spatial stochastic variation but conditional on the intensity the cases are distributed independently as a HePP. Conditional on the realization of the intensity function then data under this process can be modeled via the HePP likelihood. A variety of models result from this formulation.

First, a special case is where a spatial Gaussian process $G(s)$ is assumed and $\lambda_1(s|\boldsymbol{\theta}) = exp\{\alpha_0 + G(s)\}$ where α_0 is an intercept**.** A spatial Gaussian process defines a random field which has both long range variation (trend) and short range covariation (spatial covariation or correlation). In this way both correlation and trend can be built into the intensity process. Note that the process with $log(\lambda_1(s|\boldsymbol{\theta})) = \alpha_0 + G(s)$ is often termed a log Gaussian Cox Process (LGCP). A Gaussian process has marginal distributions that are multivariate normal with mean $\boldsymbol{\mu}$ and covariance Σ. Often the covariance is specified by a specific function which measures the covariation at given distances between locations. An example of a commonly used covariance function is

the exponential where $\sigma_{ij} = \tau exp(-\phi d_{ij})$ and σ_{ij} is th *ij* th element of Σ. Note that this specification is close to that discussed 5.2 with respect to the linear logistic spatial model except that here we do not assume a transformation from point process to conditional specification of the case-control logistic model for the point labels. To directly utilize a LGCP model with a Hierarchical context would require the normalization of the overall intensity via $\Lambda(A|\boldsymbol{\theta}) = exp\{\alpha_0\}.\int \lambda_0(s) exp\{G(s)\} ds$ over the study region. This integral could be evaluated via Monte Carlo integration within a sampler if required, if the background intensity were known. It could also be approximated via the Berman-Turner method if need be with the assumption that $\int \lambda_0(s) exp\{G(s)\} ds \approx \sum_{j=1}^{N} \lambda_{0j} w_j exp\{G_j\}$.

Approximation of the $\lambda_0(s)$ with piecewise constant λ_{0j} could be achieved in a number of ways. Of course the linear logistic formulation in 5.2 avoids the complication of evaluating such integrals, or estimating $\lambda_0(s)$, but makes inference conditional in the locations observed. This still allows the use of spatial information in the modeling but it does not directly model the locations of events per se.

5.5 Cluster formulation

Cox processes can also be specified with different types of stochastic driving processes. For example, a clustered process could be conceived where $\lambda_1(s_i|\theta) = \sum_{k=1}^{K} \mu.h(d_{ik})$ where we assume that there are k=1,...,K clusters and d_{ik} is the distance from the *k* th center, and *h(.)* is a cluster distribution function.

The cluster distribution function can be a 2D density such as a bivariate normal if required, but the assumption of a density is not necessary as in the full PP model the intensity is integrated for normalization. The cluster distribution function depends on a variance or precision parameter which controls the spread of the clusters. Note that the clusters are unobserved and usually the number K is unknown as well. This kind of model may be favored when a LGCP is regarded as too regular a process for the spatial distribution of cases of disease. If irregular clustering is more common then a process where cluster centers must be estimated could be more appropriate. Special cases of this process arise with specific assumptions such as Poisson cluster processes and the Neyman-Scott process (see Møller and Waagpeteresen,2007 for details).

Shot noise process

Cluster processes which admit a random effect for each cluster center are often called shot-noise processes so that $\lambda_1(s_i \mid \theta) = \mu \sum_{k=1}^{K} exp(\gamma_k).h(d_{ik})$ where γ_k is assumed to be zero mean Gaussian distributed. Of course extension to covariate models is straightforward when a multiplicative link is assumed so that $log(\lambda_1(s_i \mid \theta)) = \mu^* + x(s_i)'\beta + log(\sum_{k=1}^{K} exp(\gamma_k).h(d_{ik}))$. This yields a hierarchical model for parameters $\beta, \gamma,$ and centers. Of course a logistic spatial model could be assumed in this case, but the appearance of centers of unknown number complicates the problem. As centers must be sampled within McMC iterations there is a need to recomputed distances from data to new sampled centers as well as sampling different values of K. Reversible jump McMC is often proposed for these situations. Lawson (2006) section 6.5.2, and 6.5.3 gives examples of the posterior estimates of risk under these types of models. This type of problem is difficult to implement in conventional software such as WinBUGS/JAGS and resort must be made to R functions such as MCMCpack or bespoke programming.

5.5.2 Cluster approximation: exceedence estimation

One common approach to clustering is to model relative disease risk and examine posterior information that provides evidence for clustering in the risk, rather than specify a parametric cluster model. One such approach is to consider a posterior sample of $\lambda_1(s_i \mid \theta)$ values and to examine how many in the sample display elevated values. Assuming that elevated risk values may suggest clustering of risk, we could formalize this approach by assuming the set of G posterior sample values $\{\lambda_1(s_i \mid \theta^g)\}_{g=1,\ldots,G}$ and declaring a threshold for the 'elevated' risk, say c, and compute a posterior functional such as

$$Q_i = Pr(\lambda_1(s_i) > c) \approx \frac{1}{G} \sum_{g=1}^{G} I(\lambda_1(s_i \mid \theta^g) > c)$$

This computation yields an estimate of the exceedence probability (Q_i) at a location and large values of this quantity suggest 'unusually' elevated risk. The choice of two different thresholds have to be made here. First *c* must be specified. Second the level of probability deemed 'unusual' must be decided. The choice of c essentially amounts to consideration of 'null' or 'no effect' value of the intensity . One choice might be *c = 1* in that this absorbs the intensity into the background. Another possibility is to assume $c = \bar{\lambda}(s)$, an average level of risk, and this would be appropriate if the intensity of the case disease was overall higher than the control (i.e. a mis-specified control disease).

The choice of threshold for Q_i may be as a conventional p-value eg $Q_i > 0.90$ or 0.95.

There is some trade-off between choice of c and probability level of Q_i. In fact different spatial effects might appear when different combinations of these are assumed. Richardson et al (2004) have suggested using a Bayesian decision rule based on a relative risk threshold of c = 1 for aggregated count data. Using this exceedence approach has the advantage of allowing clustering to appear without making strong clustering assumptions. On the other hand the exceedences can be highly dependent on the model that is assumed for $\lambda_1(s_i \mid \theta)$. In fact patterns of exceedence tend to mimic the overall model form i.e. when a spatial trend is assumed then the exceedences will be trended, while if a more heterogeneous spatial distribution is assumed then the exceedences will be heterogeneous (even for the same dataset!). This model dependence means that there must be a sensitivity analysis carried out among candidate models and prior distributions to make sure that the final inference is robust.

6 Multivariate Extensions

Often it is necessary to examine a range of diseases and their spatial distributions. When we have multi-type processes we can still examine Bayesian hierarchical models with the additional possibility of examining correlation between diseases. Define $\lambda_k(s \mid \theta) = \lambda_{k0}(s).\lambda_{k1}(s \mid \theta)$ for the intensity of the *k* th disease. Once again we can assume that the background intensity is nuisance but could be different for each disease i.e. $\lambda_{10}(s) \neq \lambda_{20}(s) \neq \lambda_{30}(s).....$ We can proceed initially by assuming that the diseases could be modeled independently and simply estimate the intensity parameters separately. However there are interesting and useful ways of incorporating relations between diseases that can be pursued.

6.1 Competing Risk Models

If you consider that each case could have arisen from one or other of the diseases of interest then it may be important to consider the relative probability of having one disease over another. Hence we could consider the discrete probability

$Pr(k = i \mid s, \theta) = \lambda_i(s \mid \theta) / \sum_{k=1}^{K} \lambda_k(s \mid \theta)$ and then the conditional likelihood as

$$L(s \mid \theta) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} [\lambda_k(s_i \mid \theta) / \sum_{l=1}^{L} \lambda_l(s_i \mid \theta)].$$

It is straightforward to fit such an unordered categorical model, especially if the background intensities are the same. When the background intensities are different, as

would more usually be the case, then they must be estimated from control diseases or further assumptions must be made. An example of this competing risk formulation is given in Lawson and Williams (2000).

6.2 Unconditional Models

If it is not possible to condition, then the joint occurrence of the diseases must be modeled jointly. If each disease can be assumed to have a conditional likelihood that is of the heterogeneous Poisson process form (i.e. either HPP or Cox process) then a joint model could assume at the data level:

$$L(s \mid \theta) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} [\lambda_k(s_i \mid \theta) / \Lambda_{Tk}(\theta)] .$$

The normalization is for each disease, but this does not preclude the use of common components of the intensity functions linking the diseases. This is again where a Bayesian approach allows for flexible modeling: by moving correlation between diseases up the hierarchy we thereby avoid having to specify complex models for the joint data likelihood. For example, we could consider a shared component model whereby we assume:

$$\lambda(s_i \mid \theta, \theta_c) = \begin{cases} \lambda_1(s_i \mid \theta_1, \theta_c) = \lambda_{01}(s_i)\lambda_{11}(s_i \mid \theta_1, \theta_c) \\ \quad . \\ \quad . \\ \lambda_K(s_i \mid \theta_K, \theta_c) = \lambda_{0K}(s_i)\lambda_{1K}(s_i \mid \theta_K, \theta_c) \end{cases}$$

and θ_c are common components or parameters whereas elements of θ are disease specific. An example of this might be

$$\lambda_{11}(s_i \mid \theta_1, \theta_c) = exp(\alpha_{10} + v_{1i} + u_{ci})$$
$$\lambda_{12}(s_i \mid \theta_2, \theta_c) = exp(\alpha_{20} + v_{2i} + u_{ci})$$
.
$$\lambda_{1k}(s_i \mid \theta_k, \theta_c) = exp(\alpha_{k0} + v_{ki} + u_{ci})$$

where

$$v_{*i} \sim N(0, \tau_{v^*}^{-1})$$
$$u_{ci} \sim N(\bar{u}_{c_{\delta_i}}, \tau_{uc}^{-1} / n_{\delta_i})$$

$\bar{u}_{c_{\delta_i}}$ is the mean of the u values in the neighborhood of ith observation: δ_i.

The neighborhood of each point can be defined using natural neighbors, as discussed above. Here the uncorrelated random effects v_{*i} are disease specific and the common effect u_{ci} is spatially correlated with a CAR model and appears in all the disease models.

Of course if one were to condition on a joint realization of cases and controls for each of a set of multiple diseases then our joint model would convert to a multivariate binary model where y_i is an integer label for the disease type at the i th location.

The probability of a case at the I th location for any disease is

$$p_i = \sum_l \lambda_{li} / \{ \sum_l \lambda_{0li} + \sum_l \lambda_{li} \}$$

where $\lambda_{li} = \lambda_{0l}(s_i)\lambda_{1l}(s_i \mid \theta)$ and $\lambda_{0li} = \lambda_{0l}(s_i)$

If the background controls are all the same then this simplifies to $p_i = \sum_l \lambda_{1li} / \{L + \sum_l \lambda_{1li}\}$.

Denote the binary case-control indicator as z_i. Conditional on there being a case at the *i* th location then the probability it is from the k th disease is $q_{ik} = \lambda_{ki} / \{ \sum_l \lambda_{0li} + \sum_l \lambda_{li} \}$,

and this simplifies also when the controls are common. Finally the joint probability of a case of disease k at the ith location is

$Pr(y_i = k \mid z_i = 1) Pr(z_i = 1) = q_{ik} p_i = \lambda_{ki} / \{ \sum_l \lambda_{0li} + \sum_l \lambda_{li} \}$. We then simply model the probability vectors (q_{ik}, p_i) at each location.

7 Space-Time Bayesian Modeling

A natural extension to spatial case event modeling is when disease cases have attached a date or time of diagnosis so that for the *i* th person we have (s_i, t_i).

In principle, it is possible to simply extend the notion of an intensity to the time dimension and to specify

$$\lambda(s,t) = \lambda_0(s,t)\lambda_1(s,t \mid \theta),$$

where the background is now time dependent and the modeled intensity (λ_1) alos contains temporal effects. The main complication arises when the temporal dimension is random (ie times of events). Then effects must be scaled by the time gap between occurrences. While there is a large range of potential topics that could be discussed

under this heading, space limitations for this review preclude any further examination. Further examples of the application of space-time models within case event survival and general modeling applications can be found in Lawson (2006) ch8, and ch 10.

8 Examples

Putative Hazard Example

As an example of a putative hazard analysis I will examine the larynx cancer data for NW England. There are 58 cases and 1036 in the superposition of cases and controls. The location of a waste product incinerator is known and the distance from this location is precomputed (dis_i). In this example we have added a simulated covariate. For all cases and controls we have simulated age from an age distribution matched to larynx cancer incidence within age strata for the UK. This is denoted age_i The case-control indicator variable is ind_i which takes values. Appendix 2 displays the code for WInBUGS for this example. The regression parameters in this log linear model are gam0, gam1, gam2, with gam1 the coefficient for distance and gam2, the coefficient for age. gam0 is the log-scale intercept. In addition there is an uncorrelated individual level random effect (v_i) and also a spatially correlated random effect (w_i). This latter effect is assumed to have a conditional CAR specification (car.normal). and is based on neighborhoods defined by Delauney triangulation.derived from the DELDIR R package. All precisions are given sd-uniform prior distributions on the range 0 to 5.

Convergence is reached with mutiple chains by 10000 iterations, A sample of 4000 was taken and yielded the following parameter estimates (Table 1):

Table 1 Posterior average parameter estimates for the larynx cancer example

parameter	Mean	SD	2.5%	97.5%
Gam0	-10.15	1.222	-12.21	-7.603
Gam1	21.45	20.58	-9.35	71.06
Gam2	0.0386	0.0141	0.0118	0.0651

Figure 3 displays the posterior sample density for gam1 and gam2 . It is clear that while the intercept is well estimated, the distance coefficient has a large range and crosses zero in the posterior sample. This suggests there is a lack of significant relationship with distance, whereas the 95% credible limits for the age effect are both positive and so

suggests a significant though weak relationship with age. Of course this model could easily be extended to include other features of the data or alternative predictor links or forms. We could, for example, include directional effects around the putative source, or offset a deprivation index available at locations or contextually via census units. We could employ a non-linear or spline link function with covariates without altering the hierarchical model construction (see for example .Crainiceanu et al, 2007).

Cluster Detection Example

In the larynx cancer example above we might be tempted to consider a general cluster detection approach to assess whether there is in fact any support for an unusual aggregation of cases at any locale within the study region. This could be achieved by employing a cluster model such as Cox process or shot noise process described above and estimation of cluster centers. An alternative is to examine the exceedences within a posterior sample from a standard trend or random effect model. Figure 4 displays the posterior expected exceedence field for the intensity ($\lambda > 1$) for the Berman Turner integral approximation method applied to the larynx cancer example with the log-linear intensity specified as in Appendix 1, including a spatial trend. Figure 5 display the posterior expected exceedence field for the same model but with the spatial trend removed and an uncorrelated random effect added ($v_i \sim N(0,\tau_v^{-1});\ \tau_v^{-1/2} \sim U(0,5)$). Contours delimiting higher levels of exceedence probability can represent areas of clustered risk. A threshold can be assumed for these p-value surface. It is noticeable however that areas of excess risk shift depending on the underlying model: The spatail trend model suggest trending of risk into the south-west whereas the frailty model suggest an area in the mid central southern area. In fact the incinerator location is (3.545,4.140) and this lies within the 0.95 contour of risk for the frailty model but lies outside the 0.95 contour for the trend model. Clearly the goodness of fit of any putative model must be addressed when considering these highly sensitive clustering indicators.

Environmental risk field misalignment

A classic problem in environmental epidemiology is the assignment of appropriate exposures to case locations when the exposure is measured at other locations. Specifically, often pollution measures are made in networks of sample sites and these measures have to be interpolated or extrapolated to the locations of cases. As case locations are fixed points and monitoring sites are fixed networks of points then this amounts to an interpolation from one grid to another. A variety of methods can be used to achieve this. For example Bayesian Kriging could be used to provide estimates and these estimates would have associated estimation errors. The R package `spBayes` can be used for this purpose. Once these estimates are obtained they can be used in a variety of estimation procedures. However separately estimating these exposures can

lead to estimation errors in the case event models. Usually we would assume a model where the true pollutant concentration is related to the health outcome and so the use of a joint likelihood for both health outcome and observed pollutant might be proposed. This leads to a classical measurement error (ME) model formulation. Examples exist of using a conditional spatial logistic model for case event outcomes with this type of ME model (see e.g. Kim et al., 2010). In principle, this could be applied to an unconditional point process model with intensity

$\lambda(s_i,\theta) = exp\{\alpha_0 + s(x_i^t) + ...\}$

with

$x_j \sim N(x_j^t, \tau_x^{-1}\Sigma)$ $j = 1,..,M$ measurement sites,

where $\{x_j\}$ are the observed pollutant values, and S(.) denotes a predictor link (e.g. spline) .

The interpolation to the case event locations is effected by using the predictive distribution for the pollutants. This has a known form when a Gaussian distribution is assumed for x_j (see e.g. Banerjee et al, 2004). As the true value of the pollutant also appears in the health model a joint model is usually assumed so that the true value is estimated under both models. Alternatives are to use 'plug-in' (Kriged) estimates or Berkson error for x_i^t .

9 Approximate Likelihood Inference

A number of approximations can be made to components in the likelihood of Bayesian models and these can lead to simplifications in the case event situation. One approximation is to use a fine mesh over the study region and to bin points from case and control disease into the grid cells. The resulting counts can be assumed to be Poisson distributed as the arbitrary segmentation of a Poisson process yield independent Poison distributed counts. Hence both the cases and controls would be Poisson and a superposition of the counts would lead to a relative binomial model.

For instance, if the counts of case and control in the *k* th cell is y_k^{ca}, y_k^{co} then the y_k^{ca} is binomial distributed with order $n_k = y_k^{ca} + y_k^{co}$. This allows the cell counts to be modeled via a logistic model. The relevance of this model depends on the grid cell size chosen. For other mesh approximation examples see Hossain and Lawson (2009).

Alternative approximations arise when the likelihood or posterior is replaced by an approximation. INLA (Rue et al, 2009) attempts to approximate posterior distributions by Gaussian mixtures matched to the form of the density. As far as I am aware this has not been applied to point process likelihoods but could be a potential tool for faster approximate inference.

10 Conclusions

In the above I have attempted to summarize the current state of Bayesian modeling for point event data in spatial and environmental epidemiology. While much has been focused on the derivation of likelihoods and their approximation, it is important to realize that once a suitable data model is specified then via conditional independence we can formulate hierarchical models as in the more common count data situation. The main differences are 1) the evaluation of a normalizing constant integral in likelihoods; and 2) the specification of spatial correlation for events at fixed locations (rather than areas). Solutions to the first problem lie in using integral approximation /estimation methods or conditioning and label modeling, while for the second I propose the use of either full MVN correlation prior distributions or natural neighborhoods which allow the use of Markov random field (CAR) models.

Appendix 1

WinBUGS code for zeroes trick for a heterogeneous Poisson Process Bayesian model. This will invoke a Metropolis Hastings sampler as it is an arbitrary likelihood. The model is a putative source formulation where a distance variable (d[i]) is modeled via and additive link $f_i = 1 + exp(-\beta d_i)$ and a spatial trend $\lambda_i = f_i . exp\{\beta_0 + \beta_2 x_i + \beta_3 y_i\}$. The weights (w[i]) kernel density estimate of background (den[i]) and indicator function (I[i]) are read input.

```
model{

C <- 10000     #  large enough to ensure all phi[i]'s > 0

for (i in 1:N) {

f[i]<-1.+exp(-bet1*d[i])

zeros[i] <- 0

log(lam[i])<-bet0+log(f[i])+bet2*x[i]+bet3*y[i]

log(L[i])<-I[i]*log(lam[i])-w[i]*den[i]*lam[i]

phi[i] <- -log(L[i]) + C
```

```
zeros[i] ~ dpois(phi[i])}

bet0~dnorm(0,0.001)

bet1~dnorm(0,0.001)

bet2~dnorm(0,0.001)

bet3~dnorm(0,0.001) }
```

Appendix 2

WinBUGS code for a logistic spatial model with binary outcome (ind[i]) with a Bernoulli 1st level data distribution and a logit link to the probability of being a case. Here a distance model is fitted using the additive-multiplicative link (as in Appendix 1) and a linear model with intercept, age covariate (age[i]), and two random effects (v[i] and W[i]). In this case the individual frailty effect is defined as a zero mean Gaussian prior distribution: v[i]~dnorm(0,tauv), while the correlated effect is defined as an intrinsic CAR model based on Dirichlet tile neighbor adjacencies (W[1:N]~car.normal) with adj[j] and num[i] defined from the neighborhoods of a Dirichlet tesselation of the complete superposition of cases and controls (using R package DELDIR).

```
model {
for (i in 1:N){
ind[i]~dbern(p[i])
f[i]<-(1+exp(-gam1*dis[i]))*exp(gam0+gam2*age[i]+v[i]+W[i])
logit(p[i])<-log(f[i])

v[i]~dnorm(0,tauv)
res[i]<-(ind[i]-p[i])/sqrt(p[i]*(1-p[i]))
x1[i]<-x[i]
y1[i]<-y[i]
}
 for(k in 1:sumNumNeigh)
            W[1:N] ~ car.normal(adj[],wei[],num[],tauW)
tauW<-pow(sdW,-2)
sdW~dunif(0,10)
gam0~dnorm(0,0.001)
gam1~dnorm(0,0.001)
gam2~dnorm(0,0.001)
tauv<-1/pow(sdv,2)
```

```
sdv~dunif(0,100)}
```

```
Dataset:
Available from
http://www.musc.edu/biometry/people/lawsonab/Data%20and%20Programs.html
For full ODC files contact the author directly.
```

References

Banerjee, S. Carlin, B. and Gelfand, A. (2004) *Hierarchical Modeling and analysis for spatial data* CRC press New York

Berman, M. and Turner, R. (1992) Approximating Point Process likelihoods with GLIM. *Applied Statistics*, 41, 31-38

Besag, J. York, J., Mollié, A. (1991) Bayesian Image Restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-59

Crainiceanu, C. M., Ruppert, D., Carroll, R. Adarsh,J,., Goodner, B. (2007) Spatially Adaptive Bayesian P-Splines with Heteroscedastic Errors. *Journal of Computational and Graphical Statistics*, 16, 2, 265 – 288.

Diggle, P. J. (1990) A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society A*, 153, 349-362

Diggle, P. (2003) *Statistical Analysis of Spatial Point Patterns*. 2nd Ed OUP, New York

Diggle,P.J. and Rowlingson, B. (1994) A conditional approach to point process modelling of elevated risk, *Journal of the Royal Statistical Society A*, 157, 433-440

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1,3, 515-533

Hossain, M. M. and Lawson, A. B. (2009) Approximate Methods in Bayesian Point Process Spatial Models *Computational Statistics and Data Analysis (spatial analysis*

special issue: invited paper) 53, 8, *2831-2842*

Kelsall, J. and Diggle, P (1998) Spatial variation in risk of disease: a nonparametric binary regression approach *Applied Statistics*, 47, 4, 559-573

Kim, J., Lawson, A. B. , McDermott, S., Aelion, C. M. (2010) Bayesian spatial modeling of disease risk in relation to multivariate environmental risk fields *Statistics in Medicine*, 29,1, 142-157.

Knorr-Held, L. (2000) Bayesian Modeling of inseparable space-time variation in disease risk. *Statistics in Medicine*. 19, 2555-2567

Lawson, A. B. (1992) GLIM and Normalising constant models in spatial and directional data Analysis, *Computational Statistics and Data Analysis*, 13, 331-348

Lawson, A. B. (2006) *Statistical Methods in Spatial Epidemiology* 2ed, Wiley, New York.

Lawson, A. B. (2009) *Bayesian Disease Mapping: hierarchical Modeling in Spatial Epidemiology.* CRC press, New York.

Lawson, A. B. and Banerjee, S. (2009) Bayesian Spatial Analysis ch 17 in *The Sage Handbook of Spatial Analysis* ed A. S Fotheringham and P. Rogerson, Sage, New York.

Lawson, A. B. and Williams, F. L. R. (1994) Armadale: a case study in environmental Epidemiology. *Jour. Royal Stat. Soc.,A*, **157**,2,285-298

Lawson, A. B. and Williams, F.L. R. (2000) Spatial Competing Risk Models in Disease mapping . *Statistics in Medicine*, 19,17/18,2451-2468

Preparata, F. P. and M. I. Shamos (1988). *Computational Geometry - An Introduction*. Springer, 2nd printing, corrected and expanded, 1988:

Møller, J. and Waagpeteresen, R. (2007) Modern Statistics for Spatial Point Processes. *Scandinavian Journal of Statistics*, 34, 643-684

Richardson, S., Thomson, A., Best, N.G. and Elliott, P.(2004) Interpreting posterior relative risk estimates in disease mapping studies. *Environmental Health Perspectives* 112, 1016-1025

Rue, H. Martino, S., Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society* series B, 71,2,319-392

Sibson, R. (1981). "A brief description of natural neighbor interpolation (Chapter 2)". In V. Barnett. *Interpreting Multivariate Data*. Chichester: John Wiley. pp. 21–36.

Wakefield, J. (2007) Disease Mapping and spatial regression with count data *Biostatistics*, 8,2, 158-183

Waller, and Carlin, B. (2010) Disease Mapping ch 14 in *Handbook of Spatial Statistics* eds Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P., CRC Press, New York.

Figures

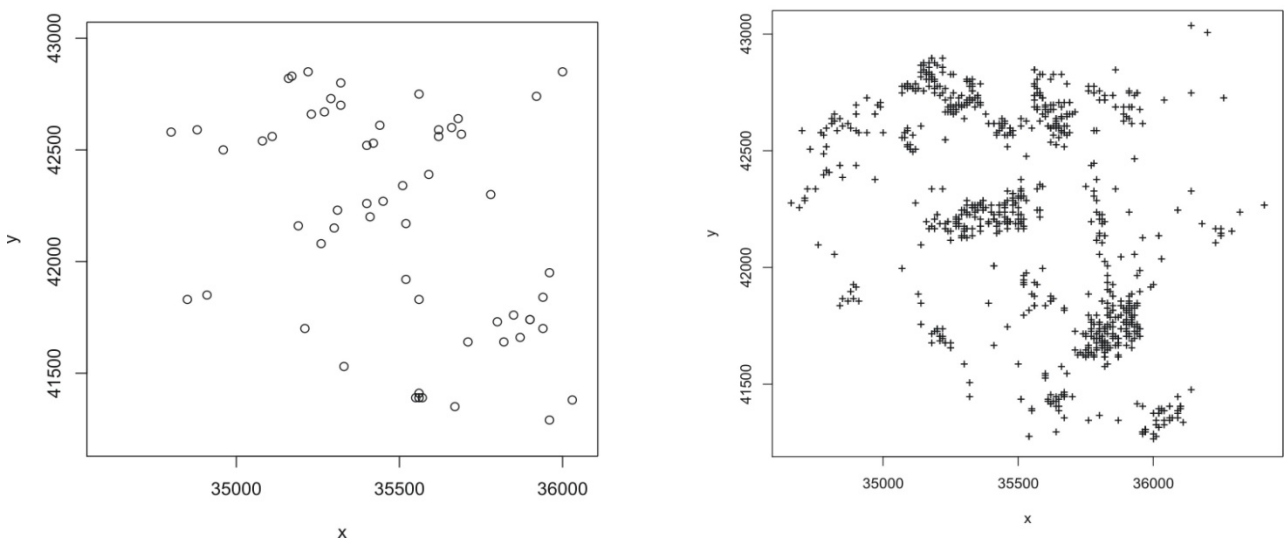


Figure 1 Larynx cancer (left panel) and respiratory cancer (right panel) incidence in a study region in Lancashire NW England

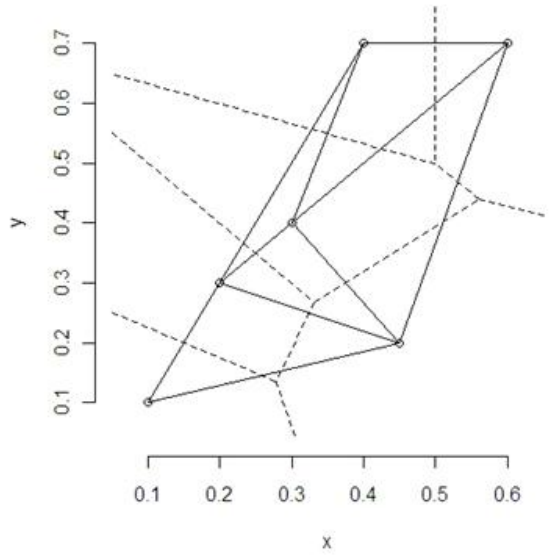


Figure 2 Tesellation/ triangulation of an arbitrary set of 6 points: Dirichlet tessellation dashed line; Delauney triangulation solid lines.

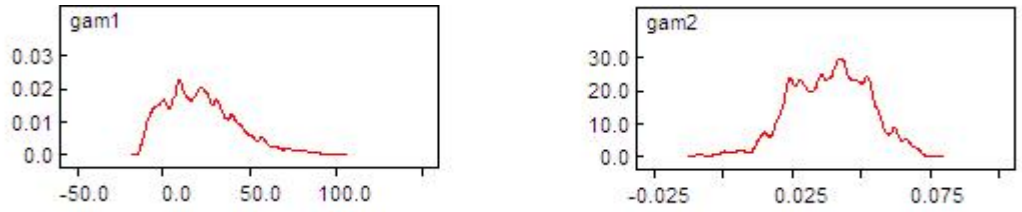


Figure 3 Posterior sample average marginal density estimates for the parameters gam1 and gam2 for the larynx cancer example.

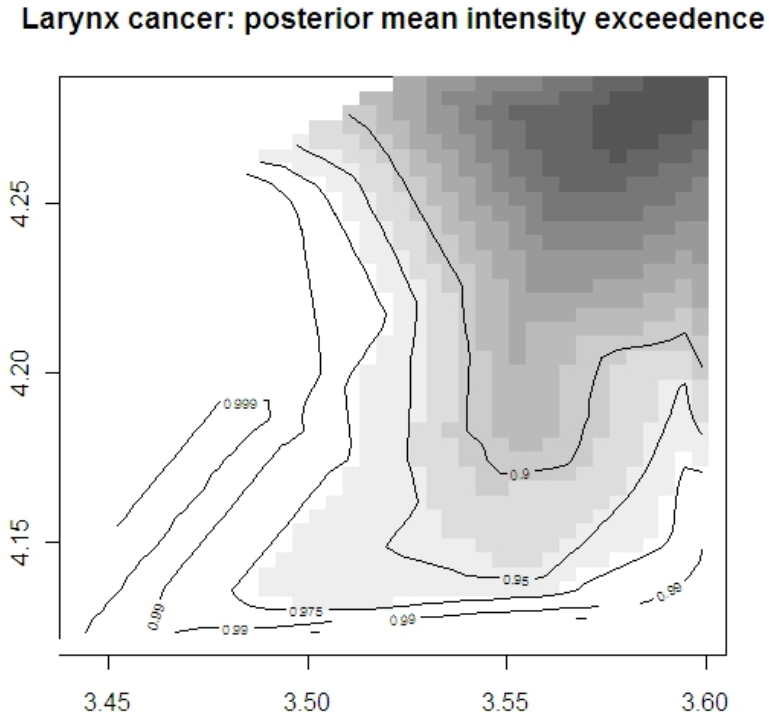


Figure 4 Average intensity exceedence probability for a BT model with spatial trend with $Pr(\lambda(s) > 1)$ estimated from posterior sample of 2000 size.

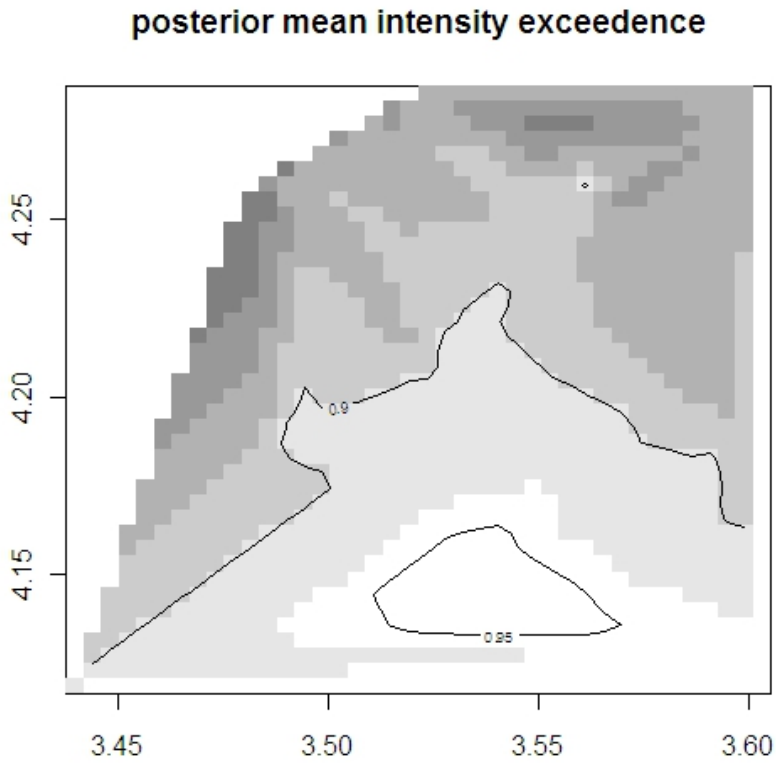


Figure 5 Average intensity exceedence probability for a BT model with no trend but with a uncorrelated frailty effect with $Pr(\lambda(s)>1)$ estimated from posterior sample of 2000 size.