ERICA CASTILHO RODRIGUES

INFERINDO A ESTRUTURA DE VIZINHANÇA EM MODELOS BAYESIANOS ESPACIAIS

Belo Horizonte

21 de fevereiro de 2011

ERICA CASTILHO RODRIGUES

ORIENTADOR: RENATO MARTINS ASSUNÇÃO

INFERINDO A ESTRUTURA DE VIZINHANÇA EM MODELOS BAYESIANOS ESPACIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Estatística.

ERICA CASTILHO RODRIGUES

Belo Horizonte

21 de fevereiro de 2011

ERICA CASTILHO RODRIGUES

Advisor: Renato Martins Assunção

INFERINDO A ESTRUTURA DE VIZINHANÇA EM MODELOS BAYESIANOS ESPACIAIS

Dissertation presented to the Graduate Program in Estatística of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Estatística.

ERICA CASTILHO RODRIGUES

Belo Horizonte

February 21, 2011

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Inferindo a estrutura de vizinhança em modelos bayesianos espaciais

ERICA CASTILHO RODRIGUES

Dissertação defendida e aprovada pela banca examinadora constituída por:

Ph. D. RENATO MARTINS ASSUNÇÃO – Orientador
Universidade Federal de Minas Gerais

Ph. D. ROSÂNGELA HELENA LOSCHI
Universidade Federal de Minas Gerais

Ph. D. MARC GENTON
Texas A&M University

Belo Horizonte, 21 de fevereiro de 2011

Acknowledgments

Agradeço primeiramente a Deus, por me permitir terminar mais essa etapa.

Não poderia deixar de agradecer em primeiro lugar aos meus pais, Ricardo e Marcia. À minha mãe, pela paciência em me ouvir todos os dias, por participar em todos os momento, sempre me confortando e me incentivando. Ao meu pai, por sempre acreditar em mim, por sempre me ajudar em tudo que fosse possível. À Gabriela, pela compreensão e pela torcida sempre constante. Gostaria de agradecer também ao Taciano, pelo amor e incentivo incondicionais. Por sempre me ouvir, me dar conselhos e por sempre compartilhar comigo cada momento dessa jornada. Não posso também deixar de agradecer às minhas grandes amigas, Poliana, Natália e Sarah, pelos ótimos momentos compartilhados e por sempre torcerem por mim.

Devo agradecer também a todos os amigos da UFMG. À Vanessa, com seu jeito meigo e generoso, quantos momentos de sufoco não passamos juntas. Às meninas do Leste, Aline, Letícia, Thaís, Márcia, Paola pelas horas de trabalho tão agradáveis ao lado de vocês. Pelos muitos momentos de descontração e por fazerem meu dia bem mais divertido. o João Vitor, pelo pouco tempo que trabalhamos juntos. Não posso esquecer da Alesssandra, ouvinte oficial do LESTE, por sempre me ajudar quando precisei. Gostaria também de agradecer a todos os mestres do Departamento de Estatística. Em especial à Rosângela e Denise, pelo incentivo sempre e pelo tanto que me ensinaram. Não posso deixar de agradecer também aos mestres do Departamento de Matemática, Bernardo e Remy, que acreditaram em mim bem no comecinho e me iniciaram na vida acadêmica. Gostaria também de agradecer ao Renato pela orientação, sempre firme, segura, por me mostrar com é importante amar o que se faz. Pela paciência em entender minhas dificuldades e por me motivar a superá-las a todo momento. Gostaria de agradecer também ao professor Marc Genton, por aceitar em participar da banca dessa dissertação e pelas sugestões dadas.

Resumo

No mapeamento de doenças, é necessário especificar uma estrutura de vizinhança para fazer inferências sobre a distribuição geográfica dos riscos relativos. Propomos um modelo em que a estrutura de vizinhança é parte do espaço paramétrico. Mantemos a propriedade de Markov usual de modelos Bayesianos espaciais: dado o grafo de vizinhança, as taxas de doença seguem um modelo auto-regressivo condicional. No entanto, o grafo em si é um parâmetro que também precisa ser estimado. Investigamos propriedades teóricas do nosso modelo. Em particular, investigamos cuidadosamente a matriz de covariância a *priori* e a *posteriori* induzida por esta estrutura de vizinhança aleatória fornecendo interpretação para cada elemento dessas matrizes. Também ilustramos as vantagens do nosso modelo com os dados simulados e exemplos de mapeamento real da doença.

Palavras-chaves: Mapeamento de doenças; Campos Aleatórios de Markov; Modelos Hierárquicos Espaciais.

Abstract

In Bayesian disease mapping, one needs to specify a neighborhood structure to make inference on the underlying geographical relative risks. We propose a model in which the neighborhood structure is part of the parameter space. We retain the Markov property of the usual Bayesian spatial models: given the neighborhood graph, the disease rates follow a conditional autoregressive model. However, the neighborhood graph itself is a parameter that also needs to be estimated. We investigate the theoretical properties of our model. In particular, we investigate carefully the prior and posterior covariance matrix induced by this random neighborhood structure providing interpretation for each element of these matrices. We also illustrate the advantages of our model with simulated data and real disease mapping examples.

Key words: Disease mapping; Markov Random Field; Spatial Hierarchical Models.

Resumo Estendido

Em modelos espaciais e, mais especificamente, em modelos para mapeamento de doenças, um ponto importante é a especificação da estrutura de vizinhança. Apesar da importância dessa escolha, em geral, são utilizadas apenas relações de adjacência entre as áreas. Existem poucas justificativas para essa prática além da facilidade de implementação através de rotinas GIS. O nosso objetivo, portanto, é propor um modelo mais flexível, no qual a própria estrutura de vizinhança faz parte do espaço paramétrico. Ela será estimada e se adaptará de acordo com a estrutura espacial da base de dados que está sendo analisada.

Para tanto, utilizamos Campos Markovianos Gaussianos e mantemos a propriedade usual de Markov dos modelos espaciais. A nossa modificação é colocar uma matriz de precisão que apresenta uma forma mais genérica e que nos permite estimar o grafo de vizinhança. Ao invés de considerarmos apenas vizinhanças locais, incluímos a possibilidade de vizinhanças com alcances mais longos. Mostramos através de exemplos simulados e com dados reais, que em muitos casos a estrutura de vizinhança local pode não ser suficiente para captar toda informação espacial contida nos dados.

Esse trabalho está organizado da seguinte maneira.

No Capítulo 1, apresentamos a notação utilizada e o modelo proposto pelos autores. Mostramos que a matriz apresentada é, de fato, uma matriz de precisão de uma distribuição normal multivariada, visto que ela é definida positiva. Encontramos ainda as distribuições condicionais do efeito aleatório de uma área dado o resto do mapa e observamos que elas apresentam um apelo intuitivo. Finalmente, ainda nessa seção, reescrevemos a distribuição conjunta de uma maneira mais simples a fim de evidenciar mais uma forma de interpretá-la e mostrar que ela consiste, na verdade, em uma mistura geométrica de distribuições normais.

No Capítulo 2, analisamos algumas propriedades teóricas do modelo. Em particular, estudamos cuidadosamente o formato das matrizes de covariância *a priori* e *a posteriori* induzidas por essa vizinhança aleatória. Apresentamos uma interpretação de cada um de seus elementos e mostramos como eles refletem o tipo de estrutura especificada através da matriz de precisão. Além disso, para um melhor entendimento do modelo, consideramos um caso específico, no o qual a inversão das matrizes de covariância *a priori* e *a posteriori* são mais simples e a interpretação dos seus termos pode ser feita de maneira mais direta.

No Capítulo 3, ilustramos algumas vantagens do modelo por meio de dados simulados e exemplos com dados reais. O primeiro exemplo considerado se refere aos dados de mortalidade infantil na cidade de Auckland. Nesse caso, notamos que a estrutura de vizinhança

mais razoável é aquela que considera todas as áreas como vizinhas umas das outras. Isso provavelmente se deve à homogeneidade da região analisada. O segundo exemplo analisado se refere à mortalidade causada por câncer pulmão, traquéia, brônquios e plêura. Novamente a estrutra de vizinhança estimada *a posteriori* foi aquela que considera todas as áreas vizinhas entre si. Para fins de comparação, consideramos uma terceira base de dados referente aos casos de mortes súbitas na infância. Calculamos algumas medidas de ajuste de modelos como DIC e o CPO e ambos mostraram que o nosso modelo apresenta resultados mais satisfatórios que os demais. Finalmente, geramos uma simulação de casos com risco constante e notamos que apesar de pontualmente todos os modelos darem a estimativa correta do risco relativo, nós fomos capazes de obter estimativas mais precisas.

Ainda no Capítulo 3, apresentamos algumas discussões a respeito das vantagens do modelo que estamos propondo. Apresentamos ainda algumas possíveis extensões que podem ser feitas, por exemplo, incluindo outros tipos de vizinhança na matriz de precisão ou considerando uma versão contínua da mesma, substituindo a soma por uma integral.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In disease mapping, the Bayesian model proposed by Besag, York and Mollié (1991), and denoted by BYM, is the most popular choice to estimate relative risks in small areas or to evaluate the effects of covariates acting as exposure measurements surrogates. Originally, BYM was introduced to model a cross-section of counts collected in a set disjoint geographical areas composing a partitioned map. Since then, BYM has been extended into several directions to include space-time generalized linear models (MacNab (2001), Beneito et al. (2008), Knorr-Held et al. (2001), Sun et al. (2000), Silva (2008)), spatial survival models (Carlin et al. (2003), Jin and Carlin (2005)), spatially-varying parameters models (Assunção et al. (2002), Assunção (2003), Gelfand et al. (2003)), and generalized additive models (Lan and Fahrmeir (2001)). Multivariate extensions incorporating two correlated sets of spatial effects have also been proposed in recent years Jin et al. (2005) , Gelfand and Vounatsou (2003), Held et al. (2005), Held et al. (2006). Many of these models can be fit using freely available software such as WinBUGS and BayesX.

BYM is based on a conditional autoregressive (CAR) model for the spatial random effects. In the CAR model, spatial dependence is expressed conditionally by requiring that the random effect in a given area, given the values in all other areas, depends only on a small set of neighboring values. More specifically, the random effect θ_i associated with the i-th area is the sum $\phi_i + \psi_i$ of two components where ϕ_i is a spatially structured random effect assigned an improper CAR prior distribution, and ψ_i is a second set of i.i.d. zero-mean normally distributed unstructured random effects. This is termed a convolution prior (Besag, York and Mollie, 1991) since the density of θ_i's will be the convolution of the joint densities of ϕ_i vector and the ψ_i vector.

An essential aspect of the BYM model and its extensions is the specification of the neighborhood structure for the areas. Although this is quite flexible and can be arbitrarily defined, in practice, it is typically based only on adjacency relationships. There are few justifications for this practice other than its conveniently easy calculation by means of GIS (Geographic information system) routines.

Notwithstanding its crucial role in the spatial Bayesian models, very few studies have considered different neighborhood structures for disease mapping problems. One notable exception is MacNab and Dean (2000) where the authors considered a model for disease rates with spatial effects structured at two geographical levels. They used infant mortality data over the period 1985-1994 from the province of British Columbia (BC) in Canada. The areas were organized in 21 health units and their sub-divisions, 79 local health areas. Health units (HUs) are administrative health divisions overseeing the functioning of the health sub-units, the local health areas (LHAs). Therefore, it was natural to expect that LHAs within the same HUs should share many health service and care characteristics beyond those determined by factors that vary smoothly in space. Hence, they assumed a random effect shared by all LHAs within the same HU. They also considered a neighborhood structure in which two LHAs are considered neighbors if they share boundaries or if there is a third LHA sharing boundaries with both local health areas. This second-order neighborhood structure is less usual and it reminds the higher autoregressive order models in the time series setting.

A more recent reference is White and Ghosh(2009), who introduced a stochastic neighborhood CAR model where the selection of the neighborhood depends on unknown parameters. They estimate how far should the neighborhood of the areas be assuming proximity weights that stay constant and equal to 1 up to a certain distance and thereafter decreases exponentially towards zero. In contrast with most of the published applied papers in disease mapping, they base their model in the proper CAR model rather than BYM. Most people prefer to use BYM, implying in an improper CAR model to deal with the spatial random effects, because the proper CAR model induces little marginal correlation between neighboring areas (See Banerjee, Carlin and Gelfand, 2004, page 81) and Assunção (2009).

These studies consider only locally larger neighborhoods than the first order neighborhood provided by the simple adjacency between the areas. Although in some situations a local neighborhood will be enough to deal with the spatial effects, we feel that spatial models should span a larger range of possibilities. Fundamentally, BYM and its variations consider the random effects being composed of either unstructured over-dispersion or small range spatial conditional variation. These are two extremes models and allowing for intermediate situations will be useful in some applications. We will show examples where the typical adjacency neighborhood structure is not sufficient to estimate the underlying risks, providing less smooth estimates than what should be inferred from the data. Our purpose is to introduce spatial effects with longer range than the immediate geographical neighborhood. This is likely to be useful specially in situations where the underlying risk changes so smoothly over larger regions as to be considered indistinguishable from a random constant value for all areas within it.

In this work, we investigate more flexible spatial conditional autoregressive models in terms of the neighborhood structure. We propose a model in which the neighborhood structure is part of the parameter space. We retain the same Markov property that is part of most Bayesian spatial models. That is, the disease rates follow a conditional autoregressive model, given the neighborhood graph. However, the neighborhood graph itself is a parameter that

also needs to be estimated. The methodology described herein permits arbitrary neighborhood extension for incorporating spatial random effects. It provides a simple mechanism for identifying the geographical extent of the conditional influence of neighboring areas.

The manuscript is organized as follows. In Section 1.2, we introduce the notation and present our model. In Section 2.1, we investigate the theoretical properties of the model. In particular, we carefully study the prior and posterior covariance matrix induced by this random neighborhood structure providing interpretation for each element of these matrices. We also present a specific and simpler case of our model allowing for a more thorough understanding of the covariance structure. In Section 3.1, we illustrate the advantages of our model with simulated data and real disease mapping examples. We end in Section 3.2 presenting the main conclusions.

1.2 Disease Mapping

A Bayesian hierarchical model is one of the main tools to make inference on the underlying relative risks of a disease observed into disjoint geographical areas of a map. These models can be described in this way: suppose that we have N geographic areas and each of them has a relative risk ψ_i for $i = 1, ..., N$, that needs to be estimated. The Bayesian inference is based on the posterior distribution of $\boldsymbol{\psi} = (\psi_1, ..., \psi_N)$ given by $f(\boldsymbol{\psi}|y_1, ..., y_N) = l(y_1, ..., y_N|\boldsymbol{\psi})f(\boldsymbol{\psi})$, where $l(y_1, ..., y_N|\boldsymbol{\psi})$ is the likelihood function and $f(\boldsymbol{\psi})$ is the prior distribution of the parameters vector $\boldsymbol{\psi}$. Conditionally on the values $\psi_1, ..., \psi_N$, the values $Y_1, ..., Y_N$ are supposed to be independent with a Poisson distribution with mean $\psi_i E_i$, where E_i is the expected value of cases under the hypotheses of constant relative risk over the areas. The modeling of the prior distribution $f(\boldsymbol{\psi})$ allows the introduction of spatial dependence between the risks such that close regions tend to have similar relative risks. This dependence appears as a Markovian structure in which the value ψ_i of one area, conditionally on all other areas values, depends only upon the ψ_j's values of its neighbors.

More specifically, the relative risk ψ_i is written as

$$\log(\psi_i) = \mu + b_i \tag{1.1}$$

where μ is the general level of the relative risk and b_i is the random effect of the i-th area. One simple possibility is to assume that the random effects b_i are independent and identically distributed with a normal distribution $N(0, \sigma^2)$. In this case, there will be no spatial effects imposed on the relative risks and the posterior distribution of ψ will reflect this independence. However, one typically anticipates a spatial dependence between the relative risks due to environmental and genetic similarities of neighboring areas. The most popular distribution to reflect the spatial structure of the data in the prior distribution was introduced by Besag et al (1991). They decomposed the random effect b_i into two parts, a non-spatially-structured component and a spatially structured component:

$$\log(\psi_i) = \mu + \theta_i + \beta_i$$

where $\theta_1, \ldots, \theta_n$ are the non-structured errors, independently and identically distributed according to a normal distribution. The random effects β_i have a spatially structured prior distribution with intrinsic CAR (ICAR) distribution. The ICAR prior distribution is an improper prior with a Markovian structure. The distribution of β_i, conditional on all the other values β_j for $j \neq i$, is given by

$$\beta_i | \beta_{-i} \sim N\left(\bar{\beta}_i, \frac{\sigma^2}{n_i}\right) \tag{1.2}$$

where $\bar{\beta}_i$ is the mean of the i-th area neighboring values β_j.

This model presents some identifiability problems of the spatial and non-spatial effects, as noticed by Eberly and Carlin(2000). To fix this problem, Leroux(1999) presented an alternative, including a parameter λ which is able to measure the effect of each component. This parameter measures the level of spatial correlation among the areas. In addition to this, it is included a parameter σ^2 to measure the random effect variance. He proposed a multivariate normal distribution for the random effects $\mathbf{b} = (b_1, \ldots, b_N)$ in (1.1) with the following precision matrix

$$\mathbf{Q} = (\sigma^2)^{-1}\left((1-\lambda)\mathbf{I} + \lambda\mathbf{R}\right) \tag{1.3}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{R}$ is the precision matrix of the ICAR model, which means that

$$\mathbf{R}_{ij} = \begin{cases} n_i & \text{if } i = j \\ -1 & \text{if } j \sim j \\ 0 & \text{otherwise} \end{cases}$$

where n_i is the number of neighbors of site i and $i \sim j$ means i neighbor of j.

For this model, the parameter λ assumes values in the interval $[0, 1]$, anda hence, the precision matrix $\mathbf{Q}$ is a weighted sum of the $\mathbf{I}$ and $\mathbf{R}$ matrices.

The BYM and Leroux models represent a mixing of two extreme situations. One situation considers a conditional dependence only on the immediate neighbors represented by the single neighborhood structure while the other situation represents the complete independence between the random effects. Both models consider that, if we have information on the immediate neighbors, no additional information about the other areas is necessary to make inference on the random effects. We think that in many practical situations this is too restrictive. Consider, for example, another extreme but possible situation in which the distribution of b_i (and hence, of ψ_i) in a given area, conditional on the rest of the map, should depend upon all the other sites, not only on the immediate neighbors. In this case, all areas are neighboring areas of all other areas. This can be a reasonable model when the region under study is small enough such that the economic, social and environmental characteristics are approximately constant over the entire region. This implies a certain exchangeability between the areas and therefore an all-inclusive dependence between the areas' pairs. Every area gives incremental additional information on a fixed area value, even if conditioning on all the other areas.

We propose a model that expands the single neighborhood structure of BYM and Leroux

models to a larger class that has geographically increasing orders of neighborhood extension. Through Bayesian updating, we can make inference about the more appropriate neighborhood structure underlying the observed data. More specifically, we extend the weighted sum precision matrix (1.3) by including matrices that represent neighborhoods of all possible orders in the simple adjacency graph.

Let each area i be a node or site of a graph and connect two nodes by one edge if they share boundaries. Let $\mathbf{A}$ be the $n \times n$ binary adjacency matrix where $\mathbf{A}_{ij} = 1$ if i and j are connected by one edge, and $\mathbf{A}_{ij} = 0$ otherwise. We say that i is a l-th order neighbor of j if the (i, j)-th element of the power matrix $\mathbf{A}^l$ is greater than zero and $\mathbf{A}_{ij}^s = 0$, for $s < l$ and $l \geq 1$. The maximum neighborhood order is given by the diameter of the graph, which is the longest path among all the shortest paths that connect two sites. In other words, it counts the minimum number of steps necessary to leave a site and go to any other site in the graph.

In our model, the vector $\mathbf{b} = (b_1, \ldots, b_N)$ in (1.1) has a multivariate normal distribution with mean zero and precision matrix given by:

$$\mathbf{Q} = (\sigma^2)^{-1} \left(\lambda_1 \mathbf{I} + \lambda_2 \mathbf{R}^{(1)} + \lambda_3 \mathbf{R}^{(2)} + \ldots + \lambda_{k+1} \mathbf{R}^{(k)} \right)$$

where $\lambda_1 + \lambda_2 + \ldots + \lambda_{k+1} = 1$ and $\lambda_i \geq 0$ for all i. The integer k is the diameter of the graph and $\mathbf{R}^{(l)}$ is the graph Laplacian that includes neighborhoods up to order l. That is,

$$\mathbf{R}_{ij}^{(l)} = \begin{cases} n_i^{(l)} & \text{if } i = j \\ -1 & \text{if } j \in \partial_i^{(l)} \\ 0 & \text{otherwise} \end{cases}$$

where $n_i^{(l)}$ is the number of neighbors of site i up to order l and $\partial_i^{(l)}$ is the set of neighbors of area i, from order 1 up to order l. Notice that, we are considering that the neighborhood relationship is symmetric, that is, $j \in \partial_i^{(l)}$ if, and only if, $i \in \partial_j^{(l)}$. It is important to point out that these matrices are linearly independent, ensuring the parameters identifiability.

This matrix is positive definite, as it satisfies the sufficient condition of being diagonal dominant. That is, for all $i = 1, \ldots, n$

$$\mathbf{Q}_{ii} > \sum_{j=1}^{N} |\mathbf{Q}_{ij}|$$

because

$$\mathbf{Q}_{ii} = \lambda_1 + \lambda_2 n_i^{(2)} + \lambda_3 n_i^{(3)} + \ldots + \lambda_{k+1} n_i^{(k)} = \lambda_1 + \sum_{j=1}^{N} |\mathbf{Q}_{ij}| > \sum_{j=1}^{N} |\mathbf{Q}_{ij}|$$

as $\lambda_1 \in (0, 1)$, and, therefore, $\mathbf{Q}$ can be a precision matrix.

From the precision matrix, it is possible to obtain the conditional distribution $b_i | \mathbf{b}_{-i}$ of each area given the vector $\mathbf{b}_{-i} = (b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n)$. It is a normal distribution with

mean $f(\mathbf{b}, \lambda)$ and variance $g(\mathbf{b}, \lambda)$ given by

$$f(\mathbf{b}, \lambda) = \frac{\lambda_2 n_i^{(1)} \bar{b}_i^{(1)} + \lambda_3 n_i^{(2)} \bar{b}_i^{(2)} + \ldots + \lambda_{k+1} n_i^{(k)} \bar{b}_i^{(k)}}{\lambda_1 + \lambda_2 n_i^{(1)} + \lambda_3 n_i^{(2)} + \ldots + \lambda_{k+1} n_i^{(k)}}$$

and

$$g(\mathbf{b}, \lambda) = \frac{\sigma^2}{\lambda_1 + \lambda_2 n_i^{(1)} + \lambda_3 n_i^{(2)} + \ldots + \lambda_{k+1} n_i^{(k)}}$$

where $\bar{b}_i^{(l)}$ is the mean of neighbors of site i up to order l. The conditional expectation is a convex linear combination of the means of its neighbors of all possible orders and the conditional variance is inversely proportional to the number of neighbors of each of these orders multiplied by their respective weight λ_l.

Let the $(n-2)$ dimensional $\mathbf{b}_{-ij}$ be the vector $\mathbf{b}$ without its $i-th$ and $j-th$ coordinates. It can be shown that the conditional correlation between the areas, $\text{Corr}(b_i, b_j | \mathbf{b}_{-ij})$ is given by

$$\text{Corr}(b_i, b_j | \mathbf{b}_{-ij}) \propto \begin{cases} \lambda_2 + \lambda_3 + \ldots + \lambda_k & \text{if } j \in \partial_i^{(1)} \\ \lambda_3 + \ldots + \lambda_k & \text{if } j \in \partial_i^{(2)} - \partial_i^{(1)} \\ . \\ . \\ \lambda_k & \text{if } j \in \partial_i^{(k)} - \bigcup_{l=1}^{k-1} \partial_i^{(l)} \end{cases} .$$

with the proportionality constant given by the inverse of

$$\sum_{l=1}^{k} \lambda_l n_i^{(l-1)} \sum_{l=1}^{k} \lambda_l n_j^{(l-1)}$$

and with $n_i^{(0)} \equiv 1$ by definition, for all $i = 1, \ldots, N$. This shows that the conditional correlation between the areas decreases with the neighborhood order l. For example, if a pair of sites are third order neighbors, the conditional correlation between them will be smaller than that between two first order neighbors. Notice also that, if all the λ_l are positive, then the conditional correlation between any pair of areas is different from zero.

We can also write the joint distribution in a more interpretable way:

$$
\begin{aligned}
f(\mathbf{b}) \quad &\propto \quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_i b_i^2(\lambda_1 + \ldots + \lambda_{k+1}n_i^{(k)}) - \lambda_2 \sum_i \sum_{j:j\in\partial_i^{(1)}} b_i b_j\right.\right. \\
&\qquad\qquad \left.\left. -\lambda_3 \sum_i \sum_{j:j\in\partial_i^{(2)}} b_i b_j - \ldots - \lambda_{k+1} \sum_i \sum_{j:j\in\partial_i^{(k)}} b_i b_j\right]\right\} \\
&= \quad \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda_1 \sum_i b_i^2 + \frac{2\lambda_2}{2}\sum_i\sum_{j:j\in\partial_i^{(1)}}b_i^2 + \ldots + \frac{2\lambda_{k+1}}{2}\sum_i\sum_{j:j\in\partial_i^{(k)}}b_i^2\right.\right. \\
&\qquad\qquad \left.\left. -\left(\frac{2\lambda_2}{2}\sum_i\sum_{j:j\in\partial_i^{(1)}}b_ib_j + \ldots + \frac{2\lambda_{k+1}}{2}\sum_i\sum_{j:j\in\partial_i^{(k)}}b_ib_j\right)\right]\right\} \\
&= \quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_i\left(\lambda_1 b_i^2 + \frac{\lambda_2}{2}\sum_{j:j\in\partial_i^{(1)}}(b_i-b_j)^2 + \ldots + \frac{\lambda_{k+1}}{2}\sum_{j:j\in\partial_i^{(k)}}(b_i-b_j)^2\right)\right]\right\} \quad (1.4)
\end{aligned}
$$

If $\lambda_l = 0$ for all $l > 1$, we are in the case of independent normal distributions. We can interpret the term associated with λ_l as a penalization for configurations showing too much variation among l-th order neighbors. The larger the value of λ_l, the smoother is the spatial pattern up to neighborhood order l.

This distribution can also be written as

$$
f(\mathbf{b}) \propto \left(\exp\left\{-\frac{1}{2\sigma^2}\sum_i b_i^2\right\}\right)^{\lambda_1}\left(\exp\left\{-\frac{1}{4\sigma^2}\sum_i\sum_{j:j\in\partial_i^{(l)}}(b_i-b_j)^2\right\}\right)^{\lambda_2}
$$
$$
\ldots \left(\exp\left\{-\frac{1}{4\sigma^2}\sum_i\sum_{j:j\in\partial_i^{(l)}}(b_i-b_j)^2\right\}\right)^{\lambda_k}
$$

which is a geometric mixture of normal distributions.

To complete the model specification, one needs to adopt prior distributions for the weights $(\lambda_1, \ldots, \lambda_k)$ and for the hyperparameter σ^2. In our applications, we assumed an inverse Gamma prior distribution for σ^2 and a uniform distribution on the k-dimensional simplex with the restriction that the $\lambda_l > 0$ and that they add to 1.

Chapter 2

Model

2.1 Model properties

To gain a better understanding of the prior and posterior distribution properties, we obtain its marginal covariance matrix in addition to the conditional correlation given earlier. To avoid a cumbersome notation and long formulas, we will consider the model that includes three different values for λ_l, one corresponding to λ_1 (associated with the individual areas and the independent case), another corresponding to λ_2 (associated with pairs of adjacent areas), and the third one, λ_3, corresponding to the highest possible order k, associated with a complete graph, where every area is neighbor of every other area. The extension to the general case is straightforward.

Considering only three components, our precision matrix reduces to

$$\mathbf{Q} = (\sigma^2)^{-1} \left(\lambda_1 \mathbf{I} + \lambda_2 \mathbf{R}^{(1)} + \lambda_3 \mathbf{R}^{(k)} \right) \tag{2.1}$$

where $\mathbf{R}^{(1)}$ is the precision matrix of the ICAR model and $\mathbf{R}^{(k)}$ is given by

$$\mathbf{R}^{(k)} = \mathrm{diag}(\mathbf{N}) - \mathbf{1}\mathbf{1}^T$$

where $\mathbf{N} = N\mathbf{1}$ with N being the total number of areas in the map and $\mathbf{1}$ denotes a N-dimensional vector of ones. The precision matrix in (2.1) can be rewritten as

$$\mathbf{Q} = (\sigma^2)^{-1} \left(\lambda_1 \mathbf{I} + \lambda_2 \mathrm{diag}(\mathbf{n}) + \lambda_3 \mathrm{diag}(\mathbf{N}) - \lambda_2 \mathbf{A} - \lambda_3 \mathbf{1}\mathbf{1}^T \right)$$

where $\mathbf{A}$ is the binary adjacency matrix and $\mathbf{A}\mathbf{1} = \mathbf{n} = (n_1, \ldots, n_N)$ is the vector which has the number of adjacent neighbors of each area.

From matrix algebra, we know that

$$\left(\mathbf{P} + \mathbf{u}\mathbf{v}^T \right)^{-1} = \mathbf{P}^{-1} - \frac{\mathbf{P}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{P}^{-1}}{1 + \mathbf{v}^T\mathbf{P}^{-1}\mathbf{u}} \tag{2.2}$$

if $\mathbf{P}$ is an invertible matrix and $\mathbf{u}$ and $\mathbf{v}$ are vectors with the same dimension. Let $\mathbf{M} =$

$\lambda_1 \mathbf{I} + \lambda_2 \text{diag}(\mathbf{n}) + \lambda_3 \text{diag}(\mathbf{N}) - \lambda_2 \mathbf{A}$ and denote by m_{ij} the ij-th element of $\mathbf{M}^{-1}$. Using the result (2.2), we have that the covariance matrix $\mathbf{Q}^{-1}$ is given by

$$
\mathbf{Q}^{-1} = \sigma^2 \left(\mathbf{M}^{-1} + \lambda_3 \frac{\mathbf{M}^{-1} \mathbf{1} \mathbf{1}^T \mathbf{M}^{-1}}{1 - \lambda_3 \mathbf{1} \mathbf{M}^{-1} \mathbf{1}^T} \right)
$$

$$
= \sigma^2 \mathbf{M}^{-1} + \frac{\sigma^2 \lambda_3}{1 - \lambda_3 \sum_{i,j} m_{ij}} \begin{bmatrix} \sum_j m_{1j} \sum_i m_{i1} & \dots & \sum_j m_{1j} \sum_i m_{iN} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \sum_j m_{Nj} \sum_i m_{i1} & \dots & \sum_j m_{Nj} \sum_i m_{iN} \end{bmatrix}
$$

As the matrix $\mathbf{M}$ is symmetric, $\mathbf{M}^{-1}$ is also symmetric and therefore, for all $l = 1, ..., N$,

$$
\sum_j m_{lj} = \sum_i m_{il} \; .
$$

Let $S_{l+} = \sum_j m_{lj} = \sum_i m_{il}$. We can write the covariance matrix as

$$
\mathbf{Q}^{-1} = \sigma^2 \mathbf{M}^{-1} + \frac{\sigma^2 \lambda_3}{1 - \lambda_3 \sum_{ij} m_{ij}} \begin{bmatrix} S_{1+} & S_{2+} & \dots & S_{N+} \end{bmatrix}^T \begin{bmatrix} S_{1+} & S_{2+} & \dots & S_{N+} \end{bmatrix} \; . \tag{2.3}
$$

To understand this covariance matrix, we consider initially the matrix $\mathbf{M}^{-1}$ by following the analytical approach adopted by Assunção and Krainski (2009). We can write

$$
\mathbf{M}^{-1} = \mathbf{M}^{-1} \left[\lambda_1 \mathbf{I} + \lambda_2 \text{diag}(\mathbf{n}) + \lambda_3 \text{diag}(\mathbf{N}) \right] \left[\lambda_1 \mathbf{I} + \lambda_2 \text{diag}(\mathbf{n}) + \lambda_3 \text{diag}(\mathbf{N}) \right]^{-1}
$$

$$
= \left[\mathbf{I} - \lambda_2 \mathbf{T} \mathbf{A} \right]^{-1} \mathbf{T}
$$

where

$$
\mathbf{T} = \text{diag} \left\{ \frac{1}{\lambda_1 + \lambda_2 n_1 + \lambda_3 N}, \dots, \frac{1}{\lambda_1 + \lambda_2 n_N + \lambda_3 N} \right\} \; . \tag{2.4}
$$

In the appendix, we show that we can expand the inverse matrix obtaining

$$
\left[\mathbf{I} - \lambda_2 \mathbf{T} \mathbf{A} \right]^{-1} \mathbf{T} = \mathbf{I} \mathbf{T} + \lambda_2 (\mathbf{T} \mathbf{A}) \mathbf{T} + \lambda_2^2 (\mathbf{T} \mathbf{A})^2 \mathbf{T} + \lambda_2^3 (\mathbf{T} \mathbf{A})^3 \mathbf{T} + \dots
$$

The elements $\left[(\mathbf{T} \mathbf{A})^l \mathbf{T} \right]_{ij}$ of the the l-th matrix in this expansion are weighted sums of all possible paths of length l starting at the i-th site and ending at the j-th site. For example, the three first matrices have elements equal to

$$
\left[(\mathbf{T} \mathbf{A}) \mathbf{T} \right]_{ij} = \frac{a_{ij}}{(\lambda_1 + \lambda_2 n_i + \lambda_3 N)(\lambda_1 + \lambda_2 n_j + \lambda_3 N)}
$$

$$
\left[(\mathbf{T} \mathbf{A})^2 \mathbf{T} \right]_{ij} = \sum_{k=1}^{N} \frac{a_{ik} a_{kj}}{(\lambda_1 + \lambda_2 n_i + \lambda_3 N)(\lambda_1 + \lambda_2 n_k + \lambda_3 N)(\lambda_1 + \lambda_2 n_j + \lambda_3 N)}
$$

$$\left[(\mathbf{TA})^3\mathbf{T}\right]_{ij} = \sum_{l=1}^{N}\sum_{k=1}^{N}\frac{a_{ik}a_{kl}a_{lj}}{(\lambda_1 + \lambda_2 n_i + \lambda_3 N)(\lambda_1 + \lambda_2 n_k + \lambda_3 N)(\lambda_1 + \lambda_2 n_l + \lambda_3 N)(\lambda_1 + \lambda_2 n_j + \lambda_3 N)}.$$

Considering the second matrix for illustration, the element $\left[(\mathbf{TA})^2\mathbf{T}\right]_{ij}$ counts all paths $i \to k \to j$ giving a weight inversely proportional to the number of immediate neighbors n_i, n_k, and n_j the areas have. Going from i to j through a highly connected area has a smaller contribution to $\mathbf{M}_{ij}^{-1}$ than if the path goes through a poorly connected intermediate area. This shows that two areas in a region of the map with highly connected areas will tend to be less correlated than two areas in a region where the areas has few immediate neighbors. Isolated

To complete the understanding of the covariance matrix $\mathbf{Q}^{-1}$ in (2.3), we consider now the value S_{i+}. We have

$$
\begin{aligned}
S_{i+} &= \sum_{j=1}^{N} m_{ij} \\
&= \sum_{j=1}^{N}\sum_{k=0}^{\infty}\lambda_2^k\left[(\mathbf{TA})^k\mathbf{T}\right]_{ij} \\
&= \sum_{k=0}^{\infty}\lambda_2^k\sum_{j=1}^{N}\left[(\mathbf{TA})^k\mathbf{T}\right]_{ij}.
\end{aligned}
$$

where we interchange the order of the terms because the sum is absolutely convergent. This quantity is a weighted sum of all paths leaving site i, the weight decreasing with the path length k. It is inversely related to the average degree of connectivity that area i has with the other areas in the graph. It is a value associated with the individual area, not with specific pairs of areas.

In summary, the covariance $\text{Cov}(b_i, b_j) = \mathbf{Q}_{ij}^{-1}$ is the sum of two components. The first one is $\mathbf{M}_{ij}^{-1}$ and represents a weighted sum of all paths from i to j with weights inversely related to their length and to the connectivity of the areas in the path. The second component is given by the product of $S_{i+}S_{+j}$ where S_{i+} is a score associated with the average connectivity of area i to the other areas in the map. The first component is influenced by the neighborhood structure through a weighted counting of each path from i to j. The second component is also influenced by the neighborhood structure but it considers only a marginal structure. Its presence in the covariance matrix position (i, j) is by means of the product of these marginal values associated with the individual areas.

2.1.1 Posterior Covariance Matrix

More relevant to the Bayesian data analysis is the posterior covariance implied by our prior spatial model. To obtain analytical expressions, assume that y_i can be approximated by a normal distribution with variance $1/\tau_y$. The posterior precision matrix is given by

$$\mathbf{Q}^* = \tau_y\mathbf{I} + \mathbf{Q} = \tau_y + \left(\sigma^2\right)^{-1}\left[\lambda_1\mathbf{I} + \lambda_2\text{diag}(\mathbf{n}) + \lambda_3\text{diag}(\mathbf{N}) - \lambda_2\mathbf{A} - \lambda_3\mathbf{11}^T\right]$$

and therefore, the covariance matrix is

$$\mathbf{Q}^{*-1} = \mathbf{M}^{*-1} + \frac{\left(\sigma^{-2}\lambda_3\right)\left(\mathbf{M}^*\right)^{-1}\left(\mathbf{1}\mathbf{1}^T\right)\left(\mathbf{M}^*\right)^{-1}}{1 - \left(\sigma^{-2}\lambda_3\right)\mathbf{1}^T\left(\mathbf{M}^*\right)^{-1}\mathbf{1}} \, .$$

where

$$\mathbf{M}^* = \left(\tau_y + \frac{\lambda_1}{\sigma^2}\right)\mathbf{I} + \frac{\lambda_2}{\sigma^2}\text{diag}(\mathbf{n}) + \frac{\lambda_3}{\sigma^2}\text{diag}(\mathbf{N}) - \frac{\lambda_2}{\sigma^2}\mathbf{A}$$

It is rather surprising that it is possible to interpret each one of the two component matrices of the covariance $\mathbf{Q}^{*-1}$. Considering initially $(\mathbf{M}^*)^{-1}$, after some algebraic manipulations analogous to those carried out earlier for the prior covariance matrix, we have that

$$\mathbf{M}^{*-1} = \left[\mathbf{I} - \left(\tau_y\lambda_3\right)\mathbf{T}^*\mathbf{A}\right]^{-1}\mathbf{T}^*$$

where

$$\mathbf{T}^* = \text{diag}\left\{\frac{1}{\tau_y + \sigma^{-2}\left(\lambda_1 + \lambda_2 n_1 + \lambda_3 N\right)}, \ldots, \frac{1}{\tau_y + \sigma^{-2}\left(\lambda_1 + \lambda_2 n_N + \lambda_3 N\right)}\right\} \, .$$

The elements of this diagonal matrix involve the data precision τ_y and the weights of the prior covariance $\sigma^{-2}\left(\lambda_1 + \lambda_2 n_N + \lambda_3 N\right)$. The relevance of each of these parts on the posterior covariance will depend on the ratio between the variance of the observations and the prior variance.

The same matrix expansion that was used earlier can be applied here. Thus we have

$$\mathbf{M}^{*-1} = \mathbf{T}^* + \left(\sigma^{-2}\lambda_2\right)\mathbf{T}^*\mathbf{A}\mathbf{T}^* + \left(\sigma^{-2}\lambda_2\right)^2\left(\mathbf{T}^*\mathbf{A}\right)^2\mathbf{T}^* + \left(\sigma^{-2}\lambda_2\right)^3\left(\mathbf{T}^*\mathbf{A}\right)^3\mathbf{T}^* + \ldots$$

As a result, the posterior covariance matrix $\mathbf{Q}^{*-1}$ has the same structure as the prior covariance matrix, being written as sum of two matrices:

$$\frac{\sigma^{-2}\lambda_3}{1 - \sigma^{-2}\lambda_3 \sum_{i,j} S_{ij}^*}\begin{bmatrix} \left(S_{1+}^*\right)^2 & S_{1+}^* S_{2+}^* & \ldots & S_{1+}^* S_{N+}^* \\ & . & & \\ & . & & \\ & . & & \\ S_{N+}^* S_{1+}^* & S_{N+}^* S_{2+}^* & \ldots & \left(S_{N+}^*\right)^2 \end{bmatrix}$$

and

$$\begin{bmatrix} m_{11}^* & m_{12}^* & \ldots & m_{1N}^* \\ . & & & \\ . & & & \\ . & & & \\ m_{N1}^* & m_{N2}^* & \ldots & m_{NN}^* \end{bmatrix} \, .$$

where m_{ij}^* is the (i,j)-th element of the matrix $\mathbf{M}^{*-1}$ and $S_{l+}^* = \sum_j m_{lj}^* = \sum_i m_{il}^*$

Therefore the posterior covariance matrix can be interpreted in the same way as the prior covariance matrix. The only difference here are the weights appearing in the counts of the

possible paths between pairs of areas. While they were equal to $(\lambda_1 + \lambda_2 n_i + \lambda_3 N)^{-1}$ for the prior, they are now equal to $\sigma^2 / (\tau_y + \sigma^2 (\lambda_1 + \lambda_2 n_i + \lambda_3 N))$. This means that, as the prior covariance, the posterior covariance can be decomposed into two components reflecting different aspects of the neighborhood graph. One component is a weighted average of all paths connecting areas i and j, longer paths having smaller weights than shorter ones. Additionally, the paths are weighted according to the connection degree of the intervening areas in the path, more connected paths having less weights. The other component of $[\mathbf{Q}^{*-1}]_{ij}$ reflects intrinsic aspects of the pair of areas i and j. It does not matter where they are located with respect to each other, this covariance component is simply a product of scores specific to each area and, in this sense, has less spatial content than the first component.

2.1.2 The specific case of two components

We consider briefly a specific case in which the inversion of the prior and posterior covariance matrices are feasible and allow an easier interpretation of the covariance matrix. Suppose that, *a priori*, the area-specific values b_i follows a multivariate normal distribution with mean zero and precision matrix

$$\mathbf{Q} = \frac{1}{\sigma^2} \left((1-\lambda)\mathbf{I} + \lambda \left(N\mathbf{I} - \mathbf{1}\mathbf{1}^T \right) \right) .$$

where $\lambda \in (0,1)$. Compared with Leroux model in (1.3), this model exchanges the first order neighborhood matrix $\mathbf{R}$ of Leroux's model by the matrix associated with the exchangeable risks model of Bernadinelli and Montomoli (1992).

Using (2.2), we can calculate the covariance matrix:

$$\mathbf{Q}^{-1} = \frac{\sigma^2}{1 - \lambda + \lambda N} \left[\mathbf{I} + \frac{\lambda}{1-\lambda} \mathbf{1}\mathbf{1}^T \right] .$$

and the correlation $\mathrm{Corr}(b_i, b_j) = \lambda$. The correlation approaches 1 as the weight of the exchangeable model increases.

We can also find the posterior covariance matrix, if we assume that the data are normally distributed with variance $(\tau_y)^{-1}$. In this case, the posterior correlation of the random effects of areas i and j is given by

$$\mathrm{Corr}(b_i, b_j | \mathbf{y}) = \frac{\lambda}{\tau_y + (\sigma^2)^{-1}(1-\lambda)}$$

This correlation is close to zero if λ is also close to zero. In the opposite direction, to get correlation close to 1, we need to have λ close to 1 and also the ratio $(\sigma^2)\tau_y$ between prior and likelihood variances close to one. That is, we need an exchangeable component with large relative weight and, at the same time, the underlying risks should have a variation similar to the likelihood variance.

Chapter 3

Application and conclusion

3.1 Some illustrative applications

In this section we will present some examples of application of our model and we will compare it whith ICAR and Leroux models.

3.1.1 Case study 1

Marshall(1990) studied the geographical variation of the infant (under 5 year of age) mortality in Auckland, New Zealand, for the period 1977-1985. Infant mortality rates differed appreciably between developed countries between 1960 and 1990, with a typical decreasing trend. However, New Zealand and Sweden presented a different pattern with increasing rates, mostly due to the increase in the sudden death syndrome rates (see Mitchell, 1990). Figure 3.1 shows in the left hand side the mortality rates on 167 geographical units of the Auckland using the population of children under 5 years of age according to the 1981 census. In the right hand side is presented a map with the median of the rates estimated by our model, which is described in more details below. The global mean is 2.633 deaths per year per thousand but there is a large variation around this average level. The spatial pattern reflects the socioeconomic differences between the areas, with those less affluent with higher rates.

We fitted our model using the software WinBUGS to obtain the posterior distribution. If we take all possible neighborhood matrices $\mathbf{R}^{(l)}$, we have l varying from 1 to 29. However, the higher order matrices are almost identical, varying by few entries from the previous order precision matrix. This shows that there will be little information to differentiate between them, which leads to severe identifiability problems. Therefore, we fit our model using the identity matrix, the neighborhood matrices up to order 11, and the matrix $\mathbf{1}\mathbf{1}^T$ full of ones. We adopted a gamma distribution with parameters equal to 0.5 and 0.005 for the inverse of σ^2 and a uniform distribution on the 13-dimensional simplex for the weights $(\lambda_1, \ldots, \lambda_{13})$. This hyperprior distributions will be used in all the following examples.

The posterior distribution of the λ_l parameters is represented in the Figure 3.2. In the right left hand side we show the estimated posterior density of λ_1, the weight parameter associated with the identity matrix and hence with the independent effects component. In the right
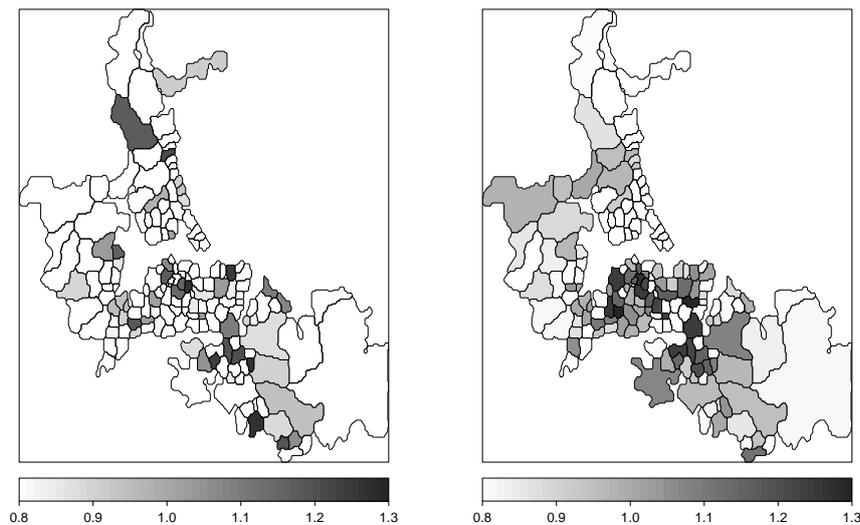
Figure 3.1: Map of SMR (left) and estimated relative risk (right) for Auckland data set.

hand side, we show all the other weights $\lambda_2, \ldots, \lambda_{13}$. The matrix with the largest weight is the identity, with the other neighborhood matrices receiving much smaller weights. In the right hand side, the density that stands out is that associated with the parameter λ_{13}. This means that the matrix $\mathbf{11}^T$ has a substantial weight if compared to the other neighborhood matrices and it is likely due to the homogeneity of the region. In this example, the matrix associated with the exchangeable random effect component is more relevant in the estimation than the adjacency matrix, which is typically used in these analysis.

In order to compare the models, we calculated the DIC criterion for this model with 13 components (hereafter, called the K components model) and also for the Leroux and BYM models. From Table 3.1 we see that, in spite of the fact that this value for our model was a bit larger then the others, the three models compared are very similar in this regard.

To see if we can retain only the exchangeable component for this data set and throw away the spatial component, we also fit two specific cases of the general model. The first one is the model that was analyzed in Section 2.1 (three components) and is composed by the spatial and exchangeable matrices. The second one is the one analyzed in Section 2.1.2 (two components), which has only the exchangeable part. The results presented in Table 3.1 show that the DIC criterion is a bit larger for the model with only two components if compared with the model composed by three. This means that the spatial effect is relevant here. Even the results showing that the exchangeable component is the most important one, it is clear, from the map of SMR (Figure 3.1), that there is an spatial structure in this data set and this behavior was confirmed here comparing these two models.

3.1.2 Case study 2

As a second example of the practical relevance of our model, we considered data from the National Cancer Institute. We analyzed the mortality rate of white women due to lung, tra-

Table 3.1: DIC criterion for Auckland data base.

K components	Two components	Three components	Leroux	BYM
837.90	847.66	846.53	831.30	832.85
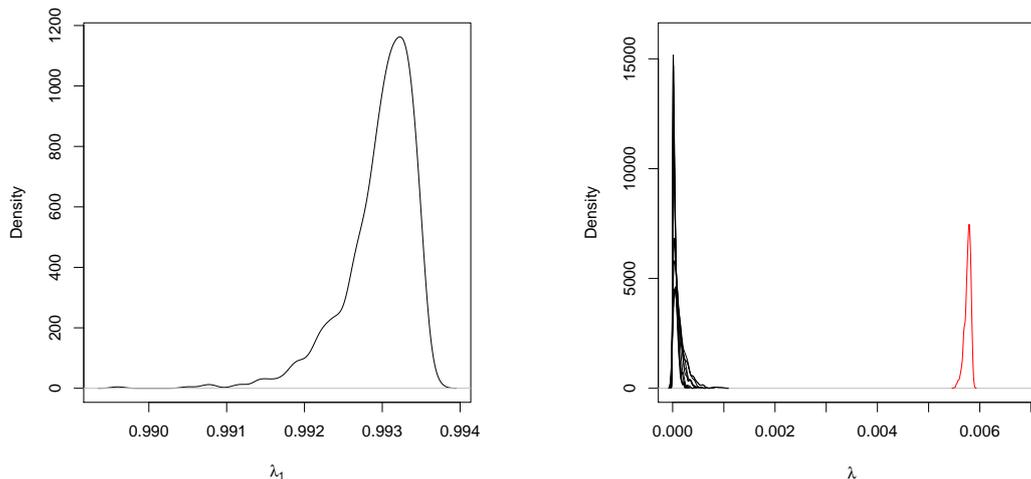
Figure 3.2: Posterior Densities of the parameters $\lambda_1, \ldots, \lambda_{14}$ for the infant mortality rates in Auckland.

chea, bronchus, and pleura cancers in the state of Tennessee. The period that was considered ranged from 1970 to 1994. It is well known that the changing in mortality patterns for this kind of cancer generally coincide with regional trends in cigarette smoking. Other studies have shown that the incidences rate is also related to occupational health hazard risks due to exposure to asbestos particles.

The map is divided into 95 counties and for each one of them we had the observed number of deaths as well as the expected number of deaths under the hypothesis of constant relative risk based on US rates. Figure 3.3 shows above the map with the relative risks estimated by the simple standardized mortality rate. Below it shows the estimates using our model. In this example, we are able to fit the model with all possible neighborhood matrices. Figure 3.4 presents the posterior densities of the weights. The graph on the left hand side shows that again the identity matrix is the most important. In the right hand side, we show density estimates of all remaining weights. The density that stands out is the weight associated with the matrix full of ones. Therefore, as happened with the previous example, a higher neighborhood order is required in the fit in order to estimate the underlying risks.

We also calculated for this data set some measures of model evaluation. For the sake of comparison, this measure was calculated for Leroux and BYM models as well. The first one that was computed is the DIC criterion and its values are presented in Table 3.2. As it was done before, besides the three main models, we also consider two additional ones: the three

Table 3.2: DIC criterion for Tennessee data base.

K components	Two components	Three components	Leroux	BYM
868.813	864.63	864.70	860.44	858.80

and the two components models. We notice then that, although our model has shown a worse performance in relation to this criterion, the difference among them is small. An interesting fact to note here is that the DIC criterion for the model of three components does not differ too much from that of the two components. This means that for this data set we could eliminate the spatial component and keep only the exchangeable one. This characteristic can be confirmed in the SMR map shown in Figure 3.3. This map shows that this disease, in fact, does not have a spatial structure. The observed number of cases is high all over the state with little spatial variation. What we see as spatial structure in the data could as well be assigned to unstructured random variation.
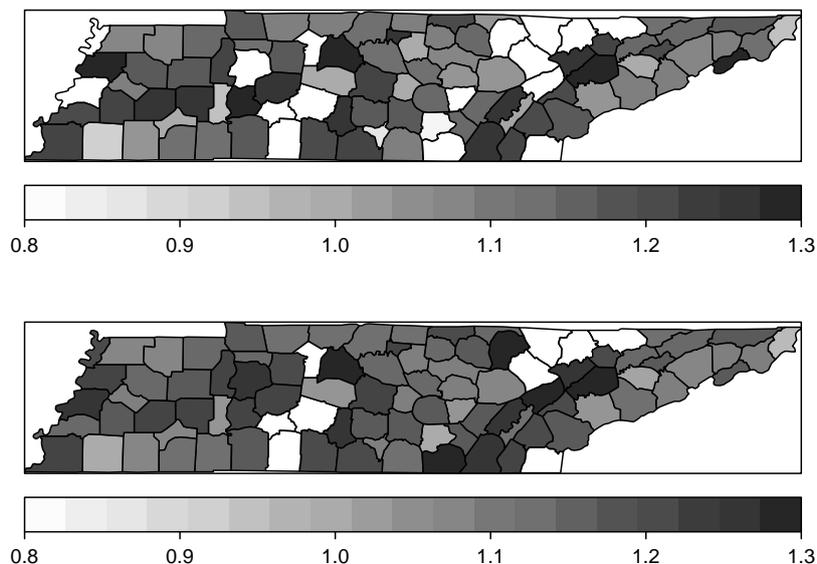
Figure 3.3: Map of SMR (above) and estimated relative risk (below).

As an alternative way of assessing the fit of the model we also calculate the conditional probability ordinate (CPO). This quantity measures the probability that we predict correctly one observation that was dropped from the data set. This is a cross validation procedure. As we had numerical problems to execute this computation considering that the data has a Poisson distribution, we consider for this case that the data are normally distributed. This assumption is reasonable for the database, because the observed number of cases is very high in all areas. Besides this, since the calculation is very computationally demanding, we were not able to apply it for the model with all matrices. We considered the three component, Leroux and BYM models. To extract the information contained on these results we calculated

Table 3.3: Summary measures of CPO considering Tennessee data set.

Number of times that the CPO was the minimum		
Three components	Leroux	BYM
36	27	32
Number of times that the CPO was less than 5 %.		
Three components	Leroux	BYM
12	11	12
Number of times that the CPO was minimal cutting in 5 %.		
Three components	Leroux	BYM
4	4	6

some summary measures, which are presented in Table 3.3. The first row shows the number of times that the CPO was minimum for each of the models, this is, we consider each area and ascertain which of the three models showed a lower value for the CPO. We notice then that our model had a worse performance for this criterion, because among the three models, it showed the highest frequency of low CPO's. It was also calculated the number of times that the CPO was less than 5% for each of the models. From the second row Table 3.3, we see that the three models were very similar with respect to this criterion, but Leroux model performed slightly better. Considering only the CPO's less than 5% we also found the number of times they were minimal in each of the models. The last row of Table 3.3 shows that, with respect to this criterion, our model and Leroux one were better than the BYM model. It is important to point out that this last criterion is the most reasonable, because the CPO's that are considered bad are really low. For example, if a particular model has the lowest CPO for one observation, compared with the others, but this value is 40%, this does not means problems of model fittings.
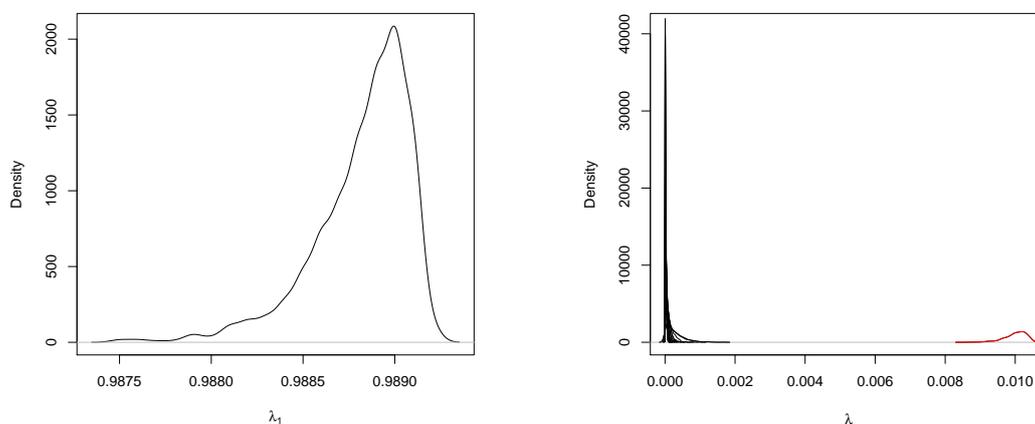
Figure 3.4: Posterior Densities of λ's.

Table 3.4: DIC criterion for North Carolina data base.

Three comopents	Leroux	BYM
469.95	471.99	471.98

3.1.3   Case study 3

We also fit the model for one classical data set of sudden infant deaths in North Carolina that was analyzed by Cressie(1991, page 386). This has also been analyzed by Kulldorff (1997) and by Lawson and Clark (2002) and it is part of many spatial statistics software manuals. Despite being intensively studied, it is surprising that we can still extract some novel information from this dataset using our model. The sudden infant death syndrome, also known as SIDS, is considered a post-neonatal death and accounts for about 7000 deaths per year in the United States. However, in spite of its impact, the cause of the syndrome is not completely understood yet. The period that was considered ranged from 1979 to 1984 and the data were registered in 100 counties of the state. We fit the model with the three matrices, as the one analyzed in Section 2.1: the identity matrix, the matrix associated with the adjacency neighborhood (ICAR matrix) and the matrix $(N-1)\mathbf{I} - \mathbf{1}\mathbf{1}^T$ corresponding to a complete graph, in which all areas are neighbors of every other area. Figure 3.5 presents the posterior densities of the weights. The first graph shows that again the identity matrix is the most important one. In the second graph, the most flat density is the one related to the matrix full of ones. Thus we see that we do not have much information about this weight, because of the shape of its density function.

In order to compare this model with Leroux and BYM ones we also fit both of them and computed the DIC criterion. These measures are presented in Table 3.4. We see that this quantity for our model is slightly smaller than the others. Therefore, our model is better in this regard.

In this case, we cannot consider that the data has normal distribution, since the disease is rare. Therefore, we computed the approximated CPO using Importance Resampling, as it was proposed by Stern and Cressie (2000). The general idea of this method is to obtain an approximated sample of a distribution that we cannot generate, by taking a resample from a realization of another one that is not too far from the one that we want. In this way, it is not necessary to pick one observation at each time in order to proceed the cross validation.

The first row of Table 3.5 shows the number of times that the CPO of each model was the least. We note that with regard to this criterion, the best performing model was Leroux one, whereas our model was very close to the BYM model. As mentioned earlier, it is interesting, however, to consider only CPO's below a certain value, because they indicate real problems in the model. In this case, we consider values below 1%. We see, from the second row of Table 3.5, that both, our model and the model presented by Leroux, were well above the BYM model with respect to this criterion. Still considering only values of CPO below 1%, we see from the the third row of the same table, that the minimum values were always of the

Table 3.5: Summary measures of CPO considering North Carolina data set.

Number of times that the CPO was the minimum.		
Three components	Leroux	BYM
47	5	43
Number of times that the CPO was less than 1%.		
Three components	Leroux	BYM
1	1	7
Number of times that the CPO was minimal cutting in 1%.		
Three components	Leroux	BYM
0	0	7

BYM model. We note, therefore, that ours and Leroux models showed a performance much higher than the BYM one in relation to these criteria.
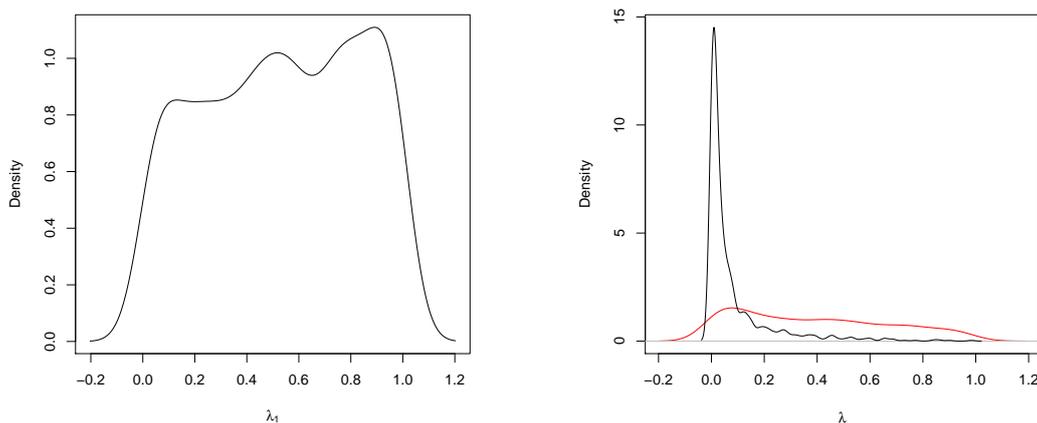
Figure 3.5: Posterior Densities of λ's for Sudden Death data set.

3.1.4   Case study 4

As another example, we analyzed the mortality rate for cancer of the trachea, bronchi and lungs in the states of São Paulo, Paraná, Santa Catarina and Rio Grande do Sul in 2007. We again fit the model whith the three matrices as in Section 2.1. The Figure show the posterior densities of the parameters. We see that the identity matrix is again the most important. However, the matrix full of ones is more important than the adjacency one. So here again the exchangeable component receives more weight than the adjacency spatial neighborhood matrix.

Comparing, in this case, the values of DIC for our model with Leroux and BYM ones we see that, as shown in Table 3.6, for this criterion our model was worse than the others.

Table 3.6: DIC criterion for Brazilian data base.

Three components	Leroux	BYM
989.07	948.48	945.85

Table 3.7: Summary measures of CPO considering Brazilian data set.

Number of times that the CPO was the minimum.		
Three components	Leroux	BYM
53	20	22
Number of times that the CPO was less than 1%.		
Three components	Leroux	BYM
1	1	2
Number of times that the CPO was minimal cutting in 1%.		
Three components	Leroux	BYM
0	0	2

We also calculated the CPO and the same summary measures considered before. We notice, from the first row of Table 3.7 that, as happened to the previous case, when we consider all the values of CPO, our model showed lower values than the others. However, the last tow rows of this table show that when we truncate the values at 1%, our model presents a performance well above the BYM model and the next to Leroux one. We realize, therefore, that in this example again, when considering the summary measure of most interest, our model performed very well.
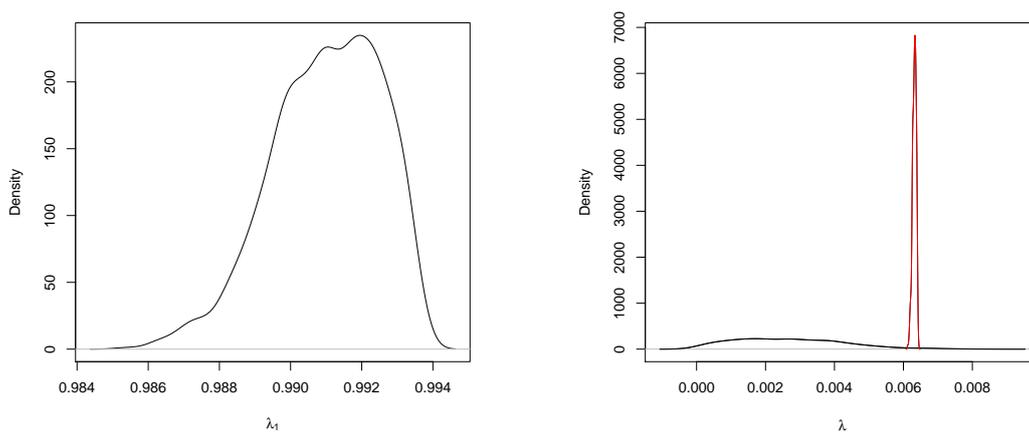
Figure 3.6: Posterior Densities of λ's for Brazilian data.

3.1.5 Simulation

As we mentioned in the description of the model, we expect it to fit quite well the data when the relative risk is approximately constant over the region. To verify this, we simulated disease counts on the Auckland map keeping the risk population equal to that of the infant mortality example. We generated conditionally independent Poisson counts with the same relative rate, equals to 2 cases for every thousand inhabitants, in all areas. We fit the model proposed by Leroux, the BYM ICAR model and our own model to this data. In our model, we used only three matrices as in Section 2.1.

The posterior densities of the relative risk for the three models are presented in Figures 3.7 and 3.8 . In all models the estimates are centered in the correct value, equal to 1. However, our model with three components has the smallest variability. The density in this case is more concentrated around the true value. This means that, by including the matrix full of ones, we were able to smooth more the estimates. The other models did not consider the possibility of conditional correlation between all areas in the map and, therefore, in this kind of situation they have less precise estimates.
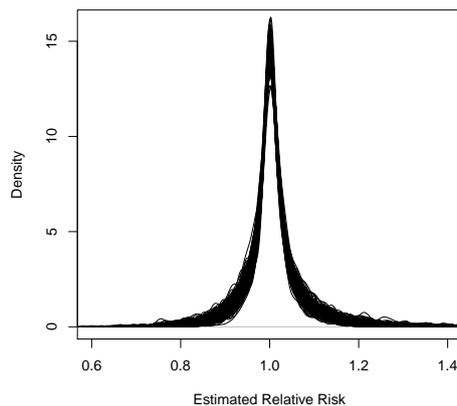

Figure 3.7: Posterior Densities of Relative Risk Bayesian estimates. Results using our model

3.2 Conclusions

In our model, we considered a precision matrix equal to a weighted average of increasing neighborhood matrices. One possibility we have not explored in this paper is to define a continuous version of this model. Let $\lambda(t)$ be a probability density function defined for $t \in [0, 1]$ and $\mathbf{R}^{(t)}$ be a continuously defined precision matrix. Assume that $\mathbf{R}^{(t)}$ as a function of t is an injective function. The precision matrix of the mixture model is given then by

$$\mathbf{Q} = \frac{1}{\sigma^2} \int_0^1 \lambda(t) \mathbf{R}^{(t)} dt \ .$$
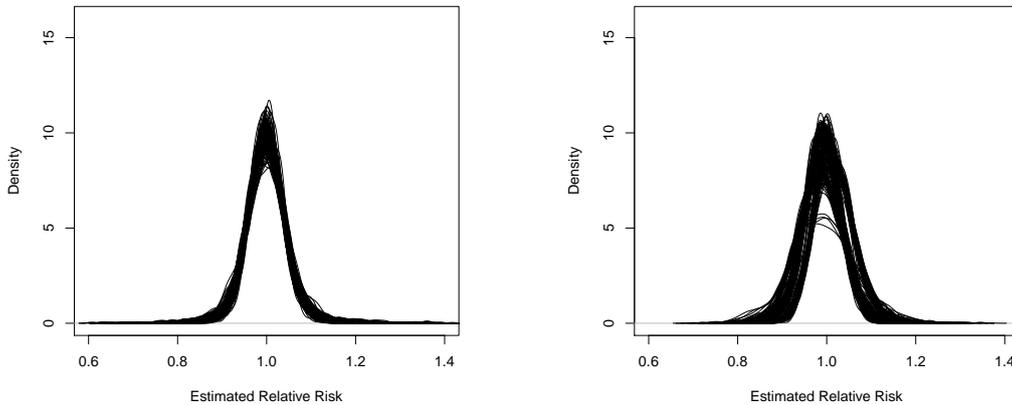
Figure 3.8: Posterior Densities of Relative Risk Bayesian estimates. Results using Leroux (left) model and the ICAR estimates (right).

This model would allow different degrees of neighborhood and could be more flexible to adapt to empirical data. Another possible extension is to treat space-time data with a precision matrix that is a mixture of simpler precision matrices.

The BYM model is very popular but one problem with it is to find the appropriate spatial smoothing degree to estimate the relative risks. In fact, other authors have noticed its tendency to oversmooth the estimates in some cases (Best *et al.*, 2005). The model we treat in this paper allows for the multiple definition of a smoothing neighborhood. In our model, the λ_j parameters control automatically this smoothing. The model can be specially useful in the situation where the underlying risk is practically constant.

We were able to interpret the conditional distributions involved in our model. The correlation between neighbors depend on the vector of λ_j values and on the graph structure.

Bibliography

[1] Assunção, R. M. and Potter, J. E. (2002). A Bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine* **14,** 2057-2075.

[2] Assunção, R. M. (2003). Space varying coefficient models for small area data. *Environmetrics* **14,** 453–473.

[3] Assunção, R. M. and Krainski, E. T. (2009). Neighborhood Dependence in Bayesian Spatial Models. *Biometrical Journal* **51,** 851–869.

[4] Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

[5] Bernardinelli, L. and Montomoli, C. (1992). Empirical bayes versus fully bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* **11,** 983–1007.

[6] Besag, J., York, J. and Mollié, A. (1991). RBayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43,** 1–20.

[7] Best, N., Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, **14,** 35–59.

[8] Brezger, A., Kneib, T., and Lang, S. (2003). BayesX-software for Bayesian inference based on Markov chain Monte Carlo simulation techniques (http://www.stat.uni-muenchen.de/ lang/).

[9] Carlin, B. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Statistics 7* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith), Oxford University Press, Oxford, 45–63.

[10] Choo, L. and Walker, G. (2008). A new approach to investigating spatial variations of disease. *Journal of the Royal Statistical Society. Series A* **171,** 395–405.

[11] Clayton, D. and Kaldor, J. (1987). Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics* **43,** 671–681.

[12] Eberly, L. E. and Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine* **19,** 2279–2294.

[13] Elliot, P. and Wartenberg, D. (2004). Spatial Epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspect* **112,** 998–1006.

[14] Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society, Series C* **50,** 201–220.

[15] Gelfand, A. E., Hyon-Jung, K., Sirmans, C.F. and Banerjee S. (2003). Spatial Modeling with Spatially Varying Coefficient Processes. *Journal of the American Statistical Association* **98,** 2057–2075.

[16] Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4,** 11–25.

[17] White G. and Ghosh S. K. (2009). A stochastic neighborhood conditional autoregressive model for spatial data. *Computational Statistics and Data Analysis* **53,** 3033–3046.

[18] Held, L., Natario, I., Fenton, S., Rue, H. and Becker, N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research* **14,** 61–82.

[19] Held, L., Graziano, G., Frank, C. and Rue, H. (2006) RJoint spatial analysis of gastrointestinal infectious diseases. *Statistical Methods in Medical Research* **15,** 465–480.

[20] Iosifescu M. *Iosifescu, Finite Markov Processes and Their Applications.* John Wiley and Sons, Chichester 1980.

[21] Iosifescu M. (1980). *Finite Markov Processes and Their Applications..* Chichester: John Wiley and Sons.

[22] Jin, X. and Carlin, B. P. (2005). Multivariate parametric spatiotemporal models for county level breast cancer survival data. *Lifetime Data Analysis* **11,** 5–27.

[23] Jin, X., Carlin, B.P. and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61,** 950–961.

[24] Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine* **17,** 2045–2060.

[25] Knorr-Held, L. and Best, N. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society, Series A* **164,** 73–85.

[26] Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics Theory and methods*, **26,** , 1481–1496.

[27] Lawosn, A.B. and Clark, A. (2002) Spatial mixture relative risk models applied to disease mapping. *Statististics in Medicine*, **21,** 359–370.

[28] Leroux, B. G., Lei, X. and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *In Statistical Models in Epidemiology; the Environment and Clinical Trials* 179–192.

[29] Lang, S. and Fahrmeir, L. (2001). Bayesian generalized additive mixed models. A simulation study. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **50,** 201–220.

[30] MacNab C. and Dean C. B. (2001). Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates. *Biometrics* **57,** 187–200.

[31] MacNab C. and Dean C. B. (2000). Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. *Statistics in Medicine* **19,** 2421–2435.

[32] Marshall, R. J. (1991). Mapping Disease and Mortality Rates Using Empirical Bayes Estimators. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **40,** 283–294.

[33] Martínez-Beneito, M. A., López-Quilez A. and Botella-Rocamora P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* **27,** 2874–2889.

[34] Silva, G. L., Dean, C. B., Niyonsenga, T. and Vanasse, A. (2008). Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine* **27,** 2381–2401.

[35] Sun, D. C., Tsutakawa, R. K., Kim, H. and He, Z. Q. (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine* **19,** 2015–2035.

[36] Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS User Manual* (Version 1.4). Cambridge: Mrc Biostatistics Unit, www.mrc-bsu.cam.ac.uk/bugs/, 2003.

[37] Spiegelhalter, D. J., Thomas, A., Best, N.G. and Lunn, D. (2003). Regression models and life tables (with discussion). *WinBUGS User Manual* (Version 1.4). Cambridge: Mrc Biostatistics Unit, www.mrc-bsu.cam.ac.uk/bugs/.

[38] Stern, H.S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine* **19,** 2377–2397.

[39] Wall, M. (2004). A close look at the spatial structure implied by the car and sar models. *Journal of Statistical Planning and Inference* **121,** 311–324.

[40] White, G. and Ghosh, S. K. (2009). A stochastic neighborhood conditional autoregressive model for spatial data. *Computational Statistics and Data Analysis* **53,** 3033–3046.

Appendix A

Demonstration of result (2.4)

A well known linear algebra result (Iosifescu (1989), page 45) states that, if $\mathbf{P}$ is a square matrix and each of the terms of the power matrix $\mathbf{P}k$ tends to zero as k increases, then the inverse $(\mathbf{I} - \mathbf{P})^{-1}$ exists and it is given by

$$(\mathbf{I} - \mathbf{P})^{-1} = \mathbf{I} + \mathbf{P} + \mathbf{P}^2 + \mathbf{P}^3 + \ldots$$

To use this result with the matrix $[\mathbf{I} - \lambda_2 \mathbf{TA}]^{-1}$, we need to show that the terms $\lambda_2^l [(\mathbf{TA})^l]_{ij}$ of the power matrix approximate zero when the power l increases. This will be done finding an upper bound. Consider initially $l = 2$. We see that

$$
\begin{aligned}
\lambda_2^2 \left[(\mathbf{TA})^2 \right]_{ij} &= \lambda_2^2 \sum_{k=1}^{N} \frac{a_{ik} a_{kj}}{(\lambda_1 + \lambda_2 n_i + \lambda_3 N)(\lambda_1 + \lambda_2 n_k + \lambda_3 N)} \\
&= \lambda_2^2 \sum_{k=1}^{N} \frac{a_{ik} a_{kj}/(n_i n_k)}{(\lambda_1/n_i + \lambda_2 + \lambda_3 N/n_i)(\lambda_1/n_k + \lambda_2 + \lambda_3 N/n_k)} \\
&< \frac{\lambda_2^2}{(\lambda_1/N + \lambda_2 + \lambda_3)^2} \sum_{k=1}^{N} \sum_{k=1}^{N} \frac{a_{ik}}{n_i} \frac{a_{kj}}{n_k} \, ,
\end{aligned}
$$

since $n_i \leq N$. As $\operatorname{diag}(1/\mathbf{n})\mathbf{A}$ is an stochastic matrix, it can be seen as a transition matrix of a random walk on the map with equal probabilities of jumping from a given area to any of its first-order neighbors. In this way, the second term in the multiplication is the probability that a random walk leaves site i and reaches site j in two steps and will be denoted by $p_{ij}^{(2)}$.

For an arbitrary $l \geq 2$, we have

$$\lambda_2^l \left[(\mathbf{TA})^l \right]_{ij} < \left(\frac{\lambda_2}{\lambda_1/N + \lambda_2 + \lambda_3} \right)^l p_{ij}^{(l)}$$

where $p_{ij}^{(l)}$ denotes the probability that the random walk goes from i to j in l steps. Therefore,

$p_{ij}^{(l)} \in [0,1]$ and since $\lambda_2/(\lambda_1/N + \lambda_2 + \lambda_3) < 1$, we have that

$$0 \leq \lim_{l \to \infty} \lambda_2^l \left[(\mathbf{TA})^{\mathbf{l}} \right]_{ij} < \lim_{l \to \infty} \left(\frac{\lambda_2}{\lambda_1/N + \lambda_2 + \lambda_3} \right)^l p_{ij}^{(l)} = 0 \,.$$

This shows that the terms of the matrix $\lambda_2^l \left[(\mathbf{TA})^{\mathbf{l}} \right]$ tends to zero as l goes to infinity and the matrix expansion is valid.

Appendix B

Another interpretation for the terms $(S_{i+}S_{+j})$

We can write $S_{i+}S_{+j}$ in a way that reflects the structure of a complete graph. To make this interpretation easier, we leave out the weights that multiply the terms of the adjacency matrix. We consider only the terms a_{ij}, which are one if i and j are neighbors and zero otherwise. Therefore we have an approximation for S_{i+} that is given by

$$S_{i+} \approx \sum_{j=1}^{N} m_{ij} = \left(\sum_{k=0}^{\infty} a_{i1}^{(k)} \right) + \left(\sum_{k=0}^{\infty} a_{i2}^{(k)} \right) + \cdots + \left(\sum_{k=0}^{\infty} a_{iN}^{(k)} \right)$$

So we see that

$$S_{i+}S_{+j} \approx \left[\left(\sum_{k=0}^{\infty} a_{i1}^{(k)} \right) + \cdots + \left(\sum_{k=0}^{\infty} a_{iN}^{(k)} \right) \right] \left[\left(\sum_{k=0}^{\infty} a_{1j}^{(k)} \right) + \cdots + \left(\sum_{k=0}^{\infty} a_{Nj}^{(k)} \right) \right]$$

$$= \underbrace{\left[\left(\sum_{k=0}^{\infty} a_{i1}^{(k)} \right) \left(\sum_{k=0}^{\infty} a_{1j}^{(k)} \right) + \cdots + \left(\sum_{k=0}^{\infty} a_{iN}^{(k)} \right) \left(\sum_{k=0}^{\infty} a_{Nj}^{(k)} \right) \right]}_{A} + \underbrace{\sum_{l \neq m} \left[\left(\sum_{k=0}^{\infty} a_{il}^{(k)} \right) \left(\sum_{k=0}^{\infty} a_{mj}^{(k)} \right) \right]}_{B}$$

Consider A first. In this sum, we have the term

$$a_{i1}^{(0)} a_{1i}^{(0)} + \cdots + a_{iN}^{(0)} a_{2N}^{(0)}$$

As $a_{ij}^{(0)} a_{1j}^{(0)} = 1$ for all j, their sum is equal to N. In A, we also have the terms

$$a_{i1}^{(0)} a_{1i}^{(1)} + \cdots + a_{iN}^{(0)} a_{2N}^{(1)}$$
$$a_{i1}^{(1)} a_{1i}^{(0)} + \cdots + a_{iN}^{(1)} a_{2N}^{(0)}.$$

Both of them count the possible ways of going from i to j in one step. If we consider the terms whose exponents summing up to 2, we have

$$a_{i1}^{(0)}a_{1i}^{(2)} + \cdots + a_{iN}^{(0)}a_{2N}^{(2)}$$
$$a_{i1}^{(1)}a_{1i}^{(1)} + \cdots + a_{iN}^{(1)}a_{2N}^{(1)}$$
$$a_{i1}^{(2)}a_{1i}^{(0)} + \cdots + a_{iN}^{(2)}a_{2N}^{(0)}$$

Each of these three terms counts the number of paths that go from i to j in 2 steps. In general, what we see is that if we take the terms whose exponent summing up to k, we will have terms of the following type:

$$a_{i1}^{(0)}a_{1i}^{(k)} + \cdots + a_{iN}^{(0)}a_{2N}^{(k)}$$
$$a_{i1}^{(1)}a_{1i}^{(k-1)} + \cdots + a_{iN}^{(1)}a_{2N}^{(k-1)}$$
$$\vdots$$
$$a_{i1}^{(k)}a_{1i}^{(0)} + \cdots + a_{iN}^{(k)}a_{2N}^{(0)}$$

This means that the paths that go from i to j in k steps are counted $k+1$ times. Therefore the term A can be written as

$$N + \sum_{n=1}^{\infty} (\text{number of paths from } i \text{ to } j \text{ in } n \text{ steps}) \, (n+1) \, .$$

Now we analyze the term B. We can see that

$$\sum_{l=1}^{N} \sum_{m \neq l} \left(\sum_{n=0}^{\infty} a_{il}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{mj}^{(n)} \right) =$$

$$\left(\sum_{n=0}^{\infty} a_{i1}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{2j}^{(n)} \right) + \left(\sum_{n=0}^{\infty} a_{i1}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{3j}^{(n)} \right) + \cdots + \left(\sum_{n=0}^{\infty} a_{i1}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{Nj}^{(n)} \right)$$

$$+ \left(\sum_{n=0}^{\infty} a_{i2}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{1j}^{(n)} \right) + \left(\sum_{n=0}^{\infty} a_{i2}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{3j}^{(n)} \right) + \cdots + \left(\sum_{n=0}^{\infty} a_{i2}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{Nj}^{(n)} \right)$$

$$\vdots$$

$$+ \left(\sum_{n=0}^{\infty} a_{iN}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{1j}^{(n)} \right) + \left(\sum_{n=0}^{\infty} a_{iN}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{2j}^{(n)} \right) + \cdots + \left(\sum_{n=0}^{\infty} a_{iN}^{(n)} \right) \left(\sum_{n=0}^{\infty} a_{(N-1)j}^{(n)} \right)$$

To see what this sum is counting, consider first the terms for which $l = 1$ and $m = 2$. In the same way that we did for the term A, we will analyze first the multiplication whose exponents sum up to 1:

$$\left[a_{i1}^{(0)} + a_{2j}^{(1)} \right] + \left[a_{i1}^{(1)} + a_{2j}^{(0)} \right]$$

This counts twice the paths that go from i to j in two steps and that have to pass through an edge connecting areas 1 and 2. If this edge indeed exists in the neighborhood graph, this is just one of its possible paths. However, if 1 and 2 are not first-order neighbors, this edge does not exist. In this case, it is as if we include this $(1, 2)$ edge in the neighborhood graph and count the additional paths that were not considered before.

Lets know consider the multiplications whose exponents summing up to 2:

$$\left[a_{i1}^{(0)} + a_{2j}^{(2)} \right] + \left[a_{i1}^{(1)} + a_{2j}^{(1)} \right] + \left[a_{i1}^{(2)} + a_{2j}^{(0)} \right]$$

We see that they represent three times the number of paths of three steps that go from i to j and pass through an edge $(1, 2)$, either truly present or not.

In general, for the terms whose exponent sum up k, we will count k times the number of possible paths of k steps from i to j and that have to pass through an edge $(1, 2)$.

We just considered here the pair of sites 1 and 2, but the sum is over all possible pairs of sites. Therefore, the term B can be written in this way

$$\sum_{k \neq l} \sum_{n=0}^{\infty} (k + 1) \, (\text{number of paths of n steps from } i \text{ to } j \text{ that pass through edge } k - l) \ .$$

This means that it counts all possible paths in the graph considering that all possible edges in the graph, in fact, exist. In other words, it counts all possible paths in a complete graph.