Stochastic Neighborhood Structure in Bayesian Spatial Models

A. Piroutek[1], R. Assunção[b], D. Duarte[c,*],

[a] *Universidade Estadual de Campinas, Instituto de Matemtica Estatstica e Cincia da Computao, Departamento de Estatstica. Cidade Universitria Zeferino Vaz Baro Geraldo 13083-970 - Campinas, SP - Brasil - Caixa-postal: 6065*

[b] *Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901. Belo Horizonte, MG, Brazil*

[c] *Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, 31270-010. Belo Horizonte, MG, Brazil.*

Abstract

Keywords: CAR model, Disease mapping, Markov Random Field, Spatial Hierarchical Models.

1. Introduction

Disease mapping has been widely used in epidemiological studies [1] and in public health surveys. This tool is often used to describe spatial variations disease incidence and prevalence, and to obtain reliable statistical estimates of local disease risk based on counts of administrative districts or regions coupled with potentially relevant background information. A brief disease mapping review can be seen in [2], [3] and [4], and applications in a variety of conditions can be seen in [5], [6] [7], [8], [9], [10], [11] and [12].

The most popular model to estimate the relative risks was proposed by Besag York and Mollié [13] denoted by BYM. This model has been studied in various statistics fields, such as survival methods ([14], [15]), multivariate models ([15], [16], [17] [18]) space-time models [19], [20], [21], [22]), and others.

*Corresponding author

Email addresses: lyne.piroutek@yahoo.com.br (A. Piroutek), assuncao@dcc.ufmg.br (R. Assunção), denise@est.ufmg.br (D. Duarte)

In BYM model, the relative risk is decomposed into two random effects: unstructured and spatially structured. The unstructured effect is an iid multivariate normal random variable. The spatially structured effect is the conditional autoregressive (CAR) model, where the spatial dependence is expressed by a Markov structure. This means that the value of the random effect of an area, given the value of all other, depends on a limited number of areas, so-called neighborhood .

An essential aspect of BYM model, similar to most of the models based on disease mapping, is the specification of the neighborhood spatial structure, denoted as neighborhood matrix W. The deterministic structure is expressed by a quite flexible matrix, commonly based on adjacency neighborhood, due to its simplicity, convenience and easily derivation through GIS routines (Geographic Information System).

An important aspect of the BYM model is the smoothing that the spatial structure causes on the relative risks estimates. When there is a association between the incidence of a disease and non-spatial factors is not convenient to use only a fixed adjacency neighborhood structure because this can cause to an oversmooth of relative risk estimates ([23], [24], [25] and [26]). We propose a solution for this problem considering a random neighborhood matrix which can be represented by a graph.

In White and Ghosh [27] the authors propose a simple extension of the CAR model, in which the neighborhood depends on the distance between areas in the map and also on the unknown parameters. For distances greater than a threshold, the entries of the neighborhood matrix follows an exponential decrease with distance. The choice of parameters is made in order to ensure two covariance matrix characteristics: positivity and sparseness.

Following a Bayesian approach, we propose a new model where for a fixed neighborhood matrix class, we associate to each matrix of this class an *a priori* distribution. The goal is to update our knowledge about these unknown matrices after observing the data. The proposed model automatically adapts the neighborhood structure according to the evidence shown in data.

We propose two *posterior* estimators for matrices in a fixed neighborhood matrix class, allowing an analysis of influence between areas. One of the proposed estimators is simple and intuitive and the other estimator is more robust, taking into account the global influence oh the areas os the map, but it is more time-consuming.

The manuscript is organized as follows. In Section 1.1 and 1.2, we introduce the notation and present some models that were proposed previously.

In section 2, we present the definition our model. In Section **??** and 3, we present three different examples of utilization and a simulation analysis of the method. In Section 4, we illustrate the use of our model for disease mapping. We end in Section 5 presenting the main conclusions.

1.1. Disease mapping

The basic idea in Disease mapping is to smooth a map of empirical relative risks by "borrowing " strength from other areas. In the following, we describe the Bayesian hierarchical model we use to model Disease mapping as usually found in literature.

We consider a region $\mathcal{D} \subset \Re^2$. The region $\mathcal{D}$ is a finite and enumerable set of geographic sites (areas) $D_1, D_2, \ldots, D_n$ where $\mathcal{D} = D_1 \cup D_2 \cup \ldots, \cup D_n$ with $D_i \cap D_j = \emptyset$ if $i \neq j$. For the sake of notation, from now on we denote the region D_i as i with $i = 1, \ldots, n$ and Y_i as the number of cases in region i.

Let N_{ij} the size of population at risk in the area i and group j, $r_j = \frac{\sum_{i=1}^{n} y_{ij}}{\sum_{i=1}^{n} N_{ij}}$ the global observed disease rate in the group j and $E_i = \sum_j N_{ij} r_j$ the expected number of cases of region i, under the hypotheses of constant relative risk over the areas. We consider that the counts $Y = (Y_1, \ldots, Y_n)$, conditional on the vector of relative risk $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)$, are random independent variables with Poisson $(\psi_i E_i)$ distribution .

We consider the relative risk $\boldsymbol{\psi}$ as random vector and the *posterior* distribution of $\boldsymbol{\psi}$, conditioned on the vector of observed cases $\boldsymbol{y} = (y_1, \ldots, y_n)$:

$$f(\boldsymbol{\psi}|\boldsymbol{y}) \propto l(y_1, \ldots, y_n) f(\boldsymbol{\psi}), \tag{1}$$

where $l(y_1, \ldots, y_n)$ is the likelihood function and $f(\boldsymbol{\psi})$ is the *a priori* distribution of parameter vector $\boldsymbol{\psi}$.

We propose an extesion of the BYM model [13] that consider a spatial dependence between the relative risks due to environmental and genetic similarities of neighboring areas. They decompose the relative risk into two components, a non-spatially structured ($\boldsymbol{\phi}$) and a spatially structured ($\boldsymbol{\theta}$):

$$\log \psi_i = \mu + \phi_i + \theta_i,$$

The non-spatial effects, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$, represent the individual contribution of each area relative risk. The elements of $\boldsymbol{\phi}$ are independents and identically distributed according to a normal distribution $N(0, 1/\tau_\phi)$ where the degree of dispersion of spatially unstructured random effects is controlled

by the unknown parameter τ_ϕ. If τ_ϕ is relatively small, the variability of random effects $(\phi_1, \ldots, \phi_n)$ is higher around their common mean of zero, meaning high relative risks variability. On the other hand, if τ_ϕ is relatively large, there is a small variation of these effects around zero..

The spatial structured effect, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ introduce the spatial dependence between the relative risks. This spatial component indicates that geographically close areas tend to present similar risks. The effect vector $\boldsymbol{\theta}$ is modeled by a Gaussian Markov Random Fields [28] called *conditional autoregressive* CAR.

$$\theta_i|(\theta_{-i}, W, \tau_\theta, \rho) \sim N\left(\frac{\rho\sum_j W_{ij}\theta_j}{\sum_j W_{ij}}, \frac{\tau_\theta^{-1}}{\sum_j W_{ij}}\right) \quad \text{where } i = 1, ..., n. \tag{2}$$

Where $\rho \in (0, 1)$, τ_θ is an hyper-parameter that is inversely proportional to the variance of θ_i, W is an $n \times n$ matrix such that $W_{ij} = W_{ji}$ and $W_{ii} = 0$. This form is most familiar as a CAR model, where W is commonly chosen as the adjacency matrix with $W_{ij} = 1$ if areas i and j are adjacent neighbors and 0 otherwise.

According to Brook Lemma [29], the joint density of the parameter vector $\boldsymbol{\theta}$ is given by:

$$\theta|(\tau_\theta, W, \rho) \sim N(0, (1 - \rho W^*)^{-1}\tau_\theta^{-1}D_{\tau_\theta}), \tag{3}$$

where D_{τ_θ} is a diagonal matrix with elements $d_{ii} = 1/n_i$ and $n_i = \sum_{j=1}^{n} W_{ij}$. If we use a neighborhood adjacency matrix, n_i is the number of neighbors of the area i. The matrix W^* is the stochastic matrix constructed from W, with $W^*_{ij} = W_{ij}/n_i$.

1.2. Maps as Graphs

Graphs are mathematical structures used to model relationships between pairs of objects of a collection and can represent the neighborhood structure of a region. We assume that each area is a node (vertice), usually located in the centroid of the area. Two areas i and j with $W_{ij} > 0$ are connected by an edge. If $W_{ij} = 0$, no edge is add. The width of the edges is proportional to the value of W_{ij}. When the matrix W is binary, only the edges different of zero are drawn. We can see in Figure 1 an example of graph constructed from a map in which we connect only adjacent areas (share borders).

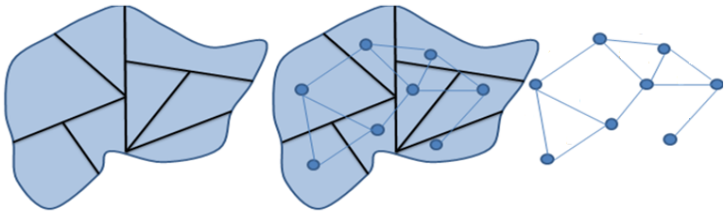


Figure 1: Representation of a graph from a map.

The map is identified by a graph $\mathcal{G}$ consisting of two sets $V_{\mathcal{G}}$ and $E_{\mathcal{G}}$, also denoted by $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, E_{\mathcal{G}})$. The first set $V_{\mathcal{G}}$ is a collection of n vertices that represents areas, and the secund set $E_{\mathcal{G}}$ is a collection of arcs or edges. Then, two vertices v e v' are connected by an edge if and only if they are neighbors, i.e., $w_{vv'} > 0$.

A subgraph of a graph $\mathcal{G}$ is a graph $\mathcal{G}'$ such that $V_{\mathcal{G}'} \subseteq V_{\mathcal{G}}$ and $E_{\mathcal{G}'} \subseteq E_{\mathcal{G}}$. A graph is called not directed if for all v and $v' \in V_{\mathcal{G}}$ the edge $(v, v') \in E_{\mathcal{G}}$ iff $(v', v) \in E_{\mathcal{G}}$.

A graph is called connected when there is a path (sequence of vertices) connecting any different pair of vertices; otherwise, the graph is called disconnected. A cycle is a path $v_1, \cdots, v_k, v_{k+1}$, onde $v_1 = v_{k+1}$, $k \geq 3$. A graph that has no cycles is called acyclic.

Given a graph $\mathcal{G}$ connected and undirected, a spanning tree $\mathcal{T}$ is an acyclic subgraph that connects all vertices. Thus, a single path in a tree connects any two nodes. Furthermore, the number of edges is equal to the number of nodes minus one. This means that, if any edge is deleted, the tree is splited into two disconnected subtrees. A single graph can be composed into more than one spanning tree.

The tree $\mathcal{T}$ is compatible with $\mathcal{G}$, if $\mathcal{T}$ can be obtained from the pruning edges of $\mathcal{G}$. We represent this definition as $\mathcal{T} \prec G$ when $\mathcal{T}$ is compatible with $\mathcal{G}$ and $\mathcal{T} \not\prec G$ otherwise.

2. Model definition

In this work, we investigate a more flexible disease-mapping model in terms of neighborhood structure. We propose the *a priori* neighborhoods matrices classes $\mathcal{W}$, allowing the consider of a stochastic neighborhood. We call this model as Stochastic Neighborhood (SN). Thus, the matrix $\mathbf{W}$ is a random object in a class $\mathcal{W}$, we can rewrite the hierarchical model as follows:

$$y_i|(\phi_i, \theta_i, E_i) \sim Poisson(E_i e^{\phi_i+\theta_i}) \quad \text{iid } i = 1, ..., n;$$

$$\phi_i|\tau_\phi \sim N(0, \tau_\phi^{-1}) \quad \text{iid } i = 1, ..., n;$$

$$\tau_\phi|(a, b) \sim Gamma(a, b);$$

$$\theta_i|(\theta_{-i}, \mathbf{W}, \tau_\theta, \rho) \sim N\left(\frac{\rho \sum_j W_{ij}\theta_j}{\sum_j W_{ij}}, \frac{\tau_\theta^{-1}}{\sum_j W_{ij}}\right) \quad i = 1, ..., n;$$

$$\tau_\theta|(c, d) \sim Gamma(c, d);$$

$$P(\mathbf{W} = W) = f(W), \quad \forall \quad W \in \mathcal{W}.$$

Then, the *a posteriori* distribution for parameters and hyper-parameters is proportional to:

$$\begin{aligned} &f(\phi_1, \ldots, \phi_n, \theta_1, \ldots, \theta_n, \tau_\phi, \tau_\theta, W|y_1, \ldots, y_n) \\ &\quad \propto \Big(\prod_{i=1}^{n} \frac{e^{-E_i e^{\phi_i+\theta_i}}(e^{\phi_i+\theta_i}E_i)^{y_i}}{y_i!}\Big) f(\phi_1, \ldots, \phi_n|\tau_\phi) f(\theta_1, \ldots, \theta_n|\tau_\theta, W) f(\tau_\phi) f(\tau_\theta) f(W) \end{aligned} \quad (4)$$

Using the shorthand notation, we denote $f(W|\phi_1, \ldots, \phi_n, \theta_1, \ldots, \theta_n, \tau_\phi, \tau_\theta, y_1, \ldots, y_n) = f(W|\gamma_{-W}, \mathbf{y})$ and the conditional distribution for $\mathbf{W}$ can be written as:

$$f(W|\gamma_{-W}, \mathbf{y}) \quad \propto \quad \det\left(2\pi \ (1 - \rho W^*)^{-1}\tau_\theta^{-1} D_{\tau_\theta}\right)^{-1/2} e^{\frac{-1}{2}\theta^t[\tau_\theta D_{\tau_\theta}^{-1}(1-\rho W^*)]\theta} f(W) \quad (5)$$

In the following, we present three examples of classes for spatial structures $\mathbf{W}$. The first class does not present any hyperparameter, the second class has one hyperparameter and the third class has both one hyperparameter and one covariate. For more examples, see appendice.

1. $\mathcal{W} = \{\mathbf{W} : \mathbf{W} \in \{\mathcal{T} : \mathcal{T} \prec \mathbf{W}^{complete}\}\}$, where $\mathcal{T}$ represents a spanning tree of graph $\mathcal{G}$ and $\mathbf{W}^{comlete}$ is a complete graph in which all areas are nearby each other. As $\mathcal{T}$ is a spanning tree, it is pruned exactly $n(n-1)/2 - (n-1)$ edges. The set of all possible spanning trees

of this class represents all realization of $\mathcal{W}_1$ and have an associated probability. If we attribute, for example, equal probability for each tree, we are in discrete uniform distribution case. Then we use the Matrix-Tree Theorem, proved by Kirchhoff [30], in order to solve the problem about the number of spanning trees of a regular graph. Two examples of spanning trees from a graph with 20 areas can be seen in figure 2.

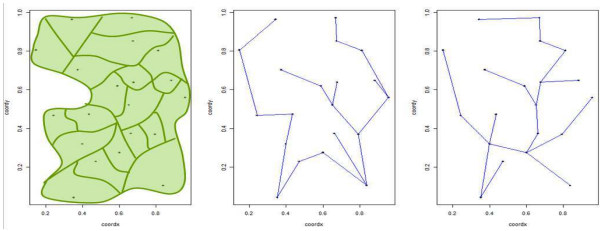


Figure 2: Two possible spanning trees in a graph with 20 areas.

2. $\mathcal{W} = \{\mathbf{W}(k) : \mathbf{W}_{ij}(k) = 1 \text{ if } j \in \{l : d_j(i) \leq d_{(k)}(i)\}, \quad \mathbf{W}_{ij}(k) = 0 \quad \text{otherwise}\}$. where $d_l(i) = ||i - l||$ with $l \neq i$, $d_{(k)}(i)$ it is k-th order statistics, and k is a hyper-parameter
This class contains only those graphs that connect the set of $k \in \{1, 2, \ldots, n-1\}$ nearest neighbors of each area. This means that this class has $n-1$ possible neighborhood matrices, and the probability of each one depends on the value k can assume. Figure 3 presents examples of this class on a map transformed into graph in which we vary the values of k.

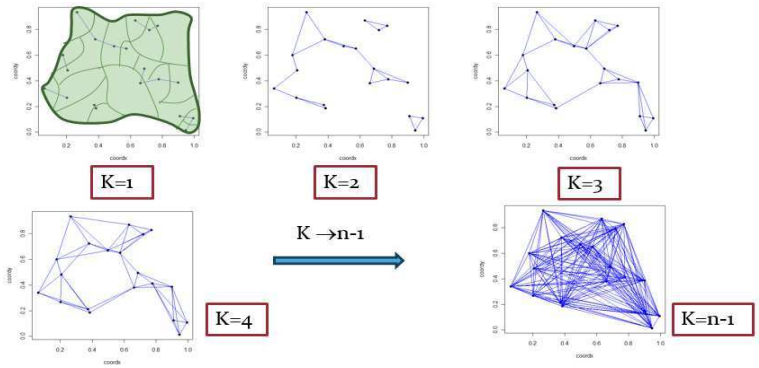


Figure 3: Example of class $\mathcal{W}_2$.

We propose two distributions for k:

- $k \sim U_d(1, (n-1))$,
- $k = 1 + k^*, \quad \text{with} \quad k^* \sim Bin((n-2), p) \quad \text{and} \quad p \in (0, 1)$,
  where U_d is the discrete uniform distribution. The hyper-parameter p varies according to the researcher's interest. If it seems more plausible few edges, i.e., if there are many connections between graphic areas, the value of p is close to zero. Otherwise, if areas are very close to each other, it is intuitive to think that each area will relate to many others around them, adding many edges in the graph. In this situation, $p \approx 1$ is more suitable.

3. $\mathcal{W} = \{\mathbf{W}(h) : W_{ij}(h) = g(x_i, x_j)\}$, where $g(x_i, x_j) = I_{[x_i \geq h]} * I_{[x_j \geq h]}$ is the product of indicator functions denoted by I and h is a hyper-parameter. As an example, we can consider the spread of H1N1 Flu Virus in 2009. Due to the traffic of people coming from many different places, cities with international airports were very affected by the virus, due to the ease spread of the disease [31], [32]. For this reason, several cities with international airports were quickly affected [33]. In this case, Mexico City and São Paulo,for instance, are considered neighbors sharing no boundary. The covariate is the gross domestic product (GDP) of each city. An example of this class can be seen in figure 4, where we adopted the covariate X as the Brazil capitals population size and an hyper-parameter $h = 2.4$millions. We can also use as covariate: size in kilometer , per capita income, Human Development Index and climate.

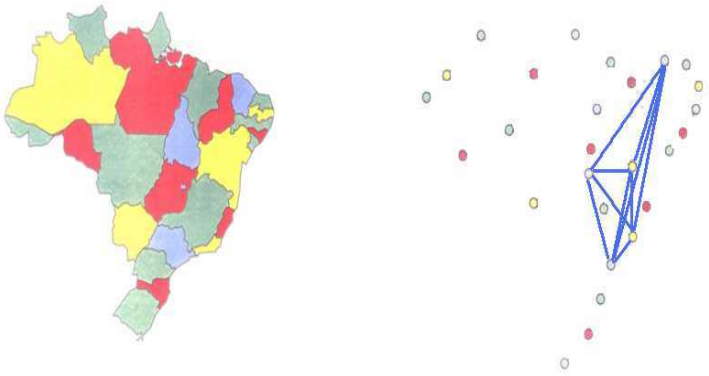


Figure 4: Example of class with covariate.

2.1. Spatial Structure Estimators

Now we are interested in estimating spatial structure given the observation $\mathbf{y}$ within a Bayesian framework. Naturally, any estimation strategy is based on the *a posteriori* distribution.

The parameters τ_ϕ e τ_θ, have a known conditional distribution and are estimated directly by the Gibbs sampler. On the other hand, the samples of parameters $\mathbf{W}$, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, are obtained through the Markov chain Monte Carlo (MCMC) algorithm. From the neighborhood matrices sample, our goal is to obtain a summary measure, such as *posteriori* mean, *posteriori* median and *posteriori* mode.

Since a the neighborhood matrix can be seen as graph (see section 1.2), we used *posteriori* estimators graphs.

In order to propose a simple and intuitive estimator we defined the first *a posteriori* estimator graph $\hat{\mathcal{G}}$ as:

$$\hat{\mathcal{G}} = argmin_{\mathcal{G}_i} \sum_{j=1}^{m} Q(\mathcal{G}_i, \mathcal{G}_j), \tag{6}$$

where the dissimilarity funtion $Q(\mathcal{G}_i, \mathcal{G}_j)$ between the graphs $\mathcal{G}_i$ and $\mathcal{G}_j$ is defined by:

$$Q(\mathcal{G}_i, \mathcal{G}_j) = \frac{\left| E_{\mathcal{G}_i} \backslash E_{\mathcal{G}_j} \bigcup E_{\mathcal{G}_j} \backslash E_{\mathcal{G}_i} \right|}{|E_{\mathcal{G}_i}| + |E_{\mathcal{G}_j}|}, \tag{7}$$

where $E_{\mathcal{G}_i}$ and $E_{\mathcal{G}_j}$ represent the sets of graph edges from $\mathcal{G}_i$ e $\mathcal{G}_j$ respectively. The notation $E_{\mathcal{G}_i} \backslash E_{\mathcal{G}_j}$ represent the difference between the two sets $E_{\mathcal{G}_i}$ and $E_{\mathcal{G}_j}$, i.e, the set of edges belonging to $\mathcal{G}_i$ and that do not belong to $\mathcal{G}_j$.

In our work, we propose one second more robust *posteriori* graph estimator $\hat{\mathcal{G}}_p$ defined by:

$$\hat{\mathcal{G}}_p = argmin_{\mathcal{G}_i} \sum_{j=1}^{m} Q_p(\mathcal{G}_i, \mathcal{G}_j), \tag{8}$$

where dissimilaritie function $Q_p(\mathcal{G}_i, \mathcal{G}_j)$ between the graphs $\mathcal{G}_i$ and $\mathcal{G}_j$ is defined by:

$$Q_p(\mathcal{G}_i, \mathcal{G}_j) = \sum_{k=1}^{n} \sum_{l=1}^{n} (A(\mathcal{G}_i)_{kl} - A(\mathcal{G}_j)_{kl})^2. \tag{9}$$

and the matrices of paths $A(\mathcal{G}_i)$ with the elements A_{kl}, $k = 1, \ldots, n$, $l = 1, \ldots, n$ is defined by:

$$A(\mathcal{G}_i)_{kl} = \text{argmin}_r \sum_{t=1}^{n} w_{kt}^{*r} w_{tl}^{*r} > 0$$

,

where W^* is the stochastic matrix constructed from the neighborhood adjacency matrix and W^{*r} is r-th power of stochastic matrix W^*. This implies that the elements of the matrix paths $A(\mathcal{G}_i)$ indicate the minimum number of required steps to go from one node to another in a graph $\mathcal{G}_i$. Hence the first order neighbors (adjacent) need only one step to become interconnect, neighbors of second order (neighbor of neighbor) need two steps, $A_{kl} = 2$, and so on.

Then, both *posteriori* graph estimators are the one that has the smallest dissimilarity between all graphs sampled via MCMC.

3. Simulation

In order to better assess the performance of the proposal method, we compare some simulation results. Our focus is to compare our results with those obtained by existing models. Specifically, we will use the BYM model, which assume the random effects spatially structured in the model CAR and neighborhood structure as adjacency. Therefore, we used the hierarchical structure described in section 2.

The observed number of cases y_i, was obtained using the bayesian spatial hierarchical model described in section 2. The rate of incidence by group j , $r_j = \frac{\sum_{i=1}^{n} y_{ij}}{\sum_{i=1}^{n} N_{ij}}$, is based on data from bronchitis and population in 127 micro-regions of Paraná, Rio Grande do Sul, Santa Catarina and São Paulo, between August 2010 and August 2011. The data are available on the Ministry of Health website Sistema de Informaes Hospitalares do SUS (SIH/SUS) and IBGE. In this case, we implemented a scenario where a determided city influences small neighboring towns. This influence is expressed through a graph that has an edge between the big city and the surrounding towns.

The selected *a priori* matrix class is the union of the class with an adjacency restriction with a new class $\mathcal{W}'$ defined as:

$$\mathcal{W}' = \begin{cases} W(k) : W_{ij} = 1 & \text{se } i \in \mathcal{B} \text{ e } j \in \{l : d_l(i) \leq d_{(k)}(i)\} \\ W(k) : W_{ij} = 0 & c.c \end{cases}$$

where $d_l(i)$ and $d_{(k)}(i)$ were definided in second example of section 2, $P(i \in \mathcal{B}) = N_i / \sum_{i=1}^{n} N_i$ and $k \sim Binomal((n-2), 0.35)$. We denote by $\mathcal{B}$ the set of cities that will be sampled as influence cities.

We also use the following *a priori* distributions:

$$\tau_\phi \sim \Gamma(0.01, 0.01);$$

$$\tau_\theta \sim \Gamma(0.01, 0.01);$$

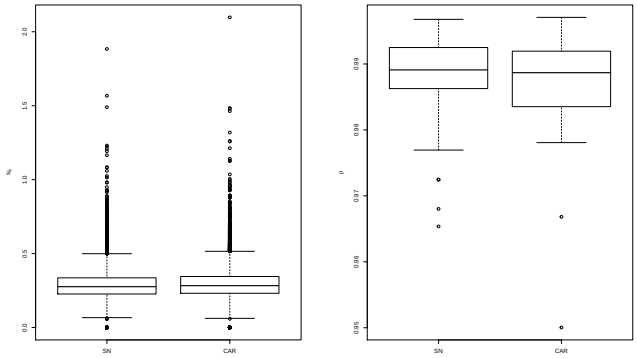
$$\rho \sim U(0, 1)$$

We ran the chains for 10000 iterations, with 500 iterations as burn-in. In each simulation, we verify the DIC value, the relative risk deviation estimative and the *a posteriori* distribution of ρ.

Figure 5(a) shows the graphics of the relative risk deviation estimative in which the boxplot of our model is in the left and the boxplot of the CAR model is in the right.

Table 1 shows that our modelpresents lower estimatives variability.

On the left side of the figure 5(b) the boxplot shows the estimates of ρ in SN model and, on the right, we see the ρ estimatives boxplot in CAR model. Both models estimates are concentrated in the true value, 0.99. However, our model shows less variability.



(a) SN model (b) CAR model

Figure 5: Box-plot of the relative risk deviation estimatives.

Finally, the table 1 shows that in 76.2% of the simulations the DIC value is smaller in our model and, in more than half of the simulations, the same behavior also occurred with the RAMSEL.

Table 1: Simulation results comparing the SN and CAR model.

Models	% smaller DIC	% smaller RMSE	% smaller RMSEL
SN	76.2	32.5	55
CAR	23.8	67.5	45

4. Application

In this section, we apply the our model for two different diseases. In the first application, we use the bronchitis data of the female population of 127 micro-regions of Paraná, Rio Grande do Sul, Santa Catarina and São Paulo, from August 2010 to August 2011. In the second application, we use data of infectious meningitis deaths in the same period of time. We choose to use two different *a priori* neighborhood matrix in each application. In the first application we choose the following class:

$$\mathcal{W} = \begin{cases} W(k) : W_{ij} = d_i(j)^{\beta} & \text{if} \quad i \in \mathcal{B} \ \text{ and } \ d_i(j) \leq 200km \\ W(k) : W_{ij} = 0 & \text{otherwise} \end{cases}$$

where $P(i \in \mathcal{B}) = \log(N_i)/\sum_{i=1}^{n} \log(N_i)$, $|\mathcal{P}| = k_1$, $\beta = (\frac{\log 0.5}{\log 50})$ e $k_1 \sim Binomal(n, 0.03)$.

We fix a negative value in β to causes a decay in value w_{ij} when the distance between the areas increases. Beside that, the β value is chosen such that $w_{ij} = 0.5$ when two cities are 50 km far. We limited the maximum value to 200 km because we believe that this threshold is reasonable when we are considering the environmental influence between areas. Since the geographical coordinates are given by latitude and longitude degrees, we use the Haversine conversion to find the distance between two areas in Kilometers.

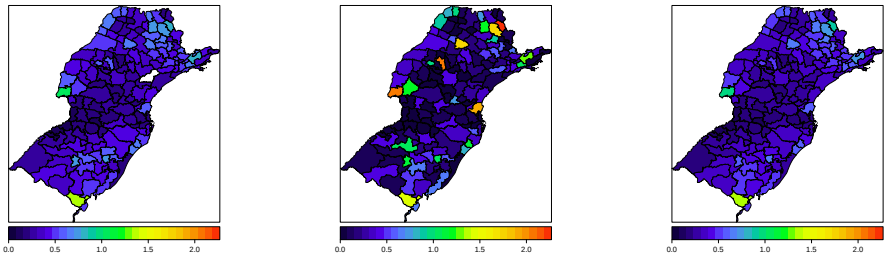
In the second application we apply the *a prior* matrix class defined in the the section 3.

From table 2, we see that our model has a much smaller DIC value than the one found by the CAR model in both applications.

Table 2: Application results of the SN and CAR models.

Models	DIC - Aplication1	DIC - Aplication2
SN	33605.29	49859.76
CAR	40130.87	60353.66

Figures 8 e 7 show the estimative relative risks maps of our model, by the SMR estimated (y_i/E_i) and the CAR model, respectively, for data bronchitis and meningitis. We conclude that our model shows smoother risk estimated value than the CAR model, because we added more edges between areas.



(a) CAR Risk (b) SMR Risk (c) SN Risk

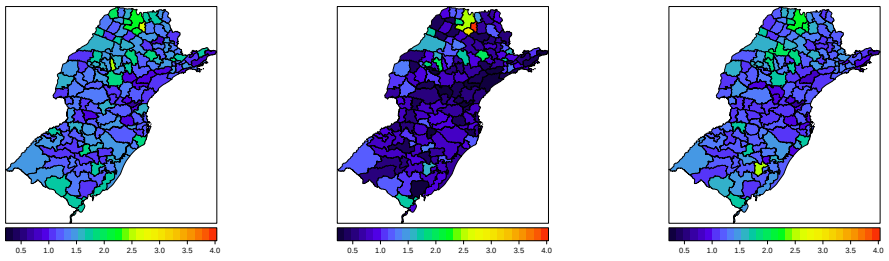
Figure 6: Risk map of bronchitis data

Figures 8(a) e 8(b) show the graphics of the deviations of the relative risks estimative and of the estimates of ρ respectively. In both figures, the boxplot of our model is the left one and boxplot of CAR model in the right side. As well as occurred with the results obtained during the simulations, our model presents lower variability of estimative.

Figure 9 and 10 present the *a posteriori* estimative on the left side. On the right side we plot graphs showing the edges that have been added to the graph when compared to adjacency graph.

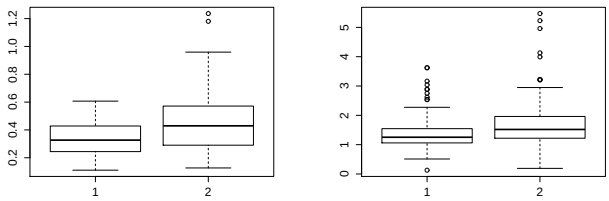
In the case of bronchitis, more edges were added near São Paulo and Campinas region when compared to the adjacency matrix. Moreover, the mean percentage of times that these edges were observed on each sample generated is 0.92 and variance is equal to 0.07. This means that, in the public health context, it is quite reasonable to analyze this region more carefully, since these cities have mutual influence.

In meningitis case, we observe the existence of four influence cities: Cu-



(a) CAR Risk (b) SMR Risk (c) SN Risk

Figure 7: Risk map of meningitis data



(a) Bronchitis data (b) Meningitis datal

Figure 8: Box-plot of the deviations of the estimated ρ. The left side represents the SN model and the right represents the CAR model.

ritiba, Porto Alegre, So Paulo and Campinas. The mean percentage of times that these edges were observed on each sample generateof edges added (in addition to the adjacency matrix) is 0.43 with variance equal to 0.42.

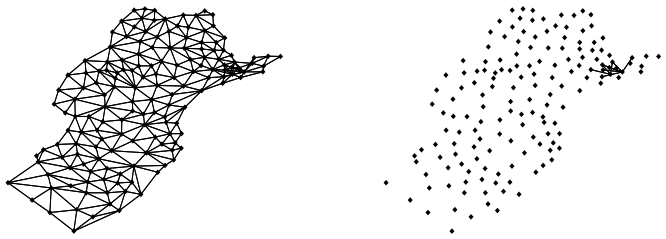


Figure 9: The left graph represents the estimated graph and the rigth graph shows the edges that have been added to the graph when compared to adjacency graph in bronchitis data.

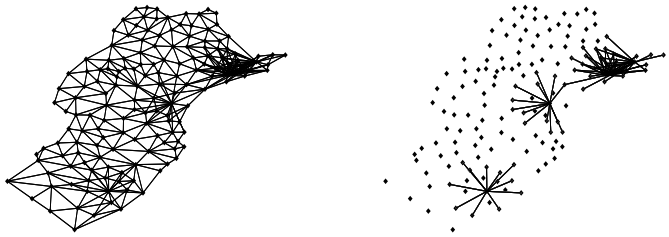


Figure 10: The left graph represents the estimated graph and the rigth graph shows the edges that have been added to the graph when compared to adjacency graph in Meningitis data.

5. Conclusion

In this work, we consider an extension of models where two random effects can explain the relative risk of a particular diseas. In the BYM model, the

neighborhood matrix is fixed before the analysis and is usually based on adjacency for convenience. We allow the spatial matrix to be a random elements of a neighbor class. We proposed classes of matrices, which allowed the usage of methods (e.g., the MCMC) to make inferences about the spatial structure. Furthermore, we proposed two *posteriori* estimators and tested our model in different situations.

The simulation lead us to conclude that the results obtained in space-structure estimative in a lattice were satisfactory. The results attested a better fit of our model when compared to CAR.

Finally, we can conclude that our proposal is more suitable for disease mapping, since it makes unnecessary to select a single neighborhood matrix and the results are more accurate.

6. Appendix

6.1. Simple Class of random matrices

In this first group we propose some simple classes that involve hyper-parameters such as distances between areas, nearest neighbor and spanning trees.

1. $\mathcal{W}_1 = \{\mathbf{W}(d) : \mathbf{W}_{ij}(d) = 1 \quad \text{if} \quad d_j(i) \leq d; \quad \mathbf{W}_{ij}(d) = 0 \quad \text{otherwise}\}$. This class also incorporates an unknown hyper-parameter of distance $d \in \mathbb{R}^+$.

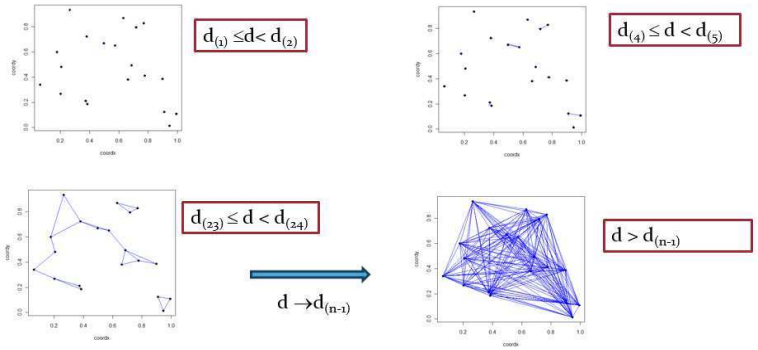


Figure 11: Example of class $\mathcal{W}_3$

We propose two distributions for d:

- $d \sim U(d_{(1)}(\bullet), d_{(n)}(\bullet))$, ie, the minimum and maximum distance between all possible pairs of vertices of the graph,
- $d = d_{(n)}(\bullet) \times d^*$, com $d^* \sim Beta(\alpha, \beta)$.
  The hyper-parameters α and β allow greater flexibility in the choice of d. If α is equal to one, we give more weight to small values of d. If we also set β equal to one, we face a case where all the possible values of p have the same weight.

2. $\mathcal{W}_2 = \{\mathbf{W}(\mathcal{C}) : \mathcal{T}_{\mathcal{AG}} \prec \mathbf{W}(\mathcal{C}^-) \prec \mathbf{W}^{adj}\}$, where $\mathbf{W}(\mathcal{C}^-)$ is the matrix compatible with the adjacency matrix $\mathbf{W}^{adj}$ in which it is obtained from the pruning of $\mathcal{C}$ edges of $\mathbf{W}^{adj}$. It can be atributed a distribution for the hyper-parameter $\mathcal{C}$.
   - $\mathcal{C} \sim U_d\Big(0, [\mathcal{C}_{max} - (n-1)]\Big)$,
     where $\mathcal{C}_{max}$ is the number of edges in graph that uses only adjacency neighborhood.
   - $\mathcal{C} \sim Binomial\Big([\mathcal{C}_{max} - (n-1)], p\Big)$.
     The hyper-parameter p allows more flexibility in choosing the number of edges $\mathcal{C}$, which are cut from the adjacency matrix $\mathbf{W}^{adj}$.
3. $\mathcal{W}_3 = \{\mathbf{W}(\mathcal{C}) : \mathcal{T}_{\mathcal{AG}} \prec \mathbf{W}(\mathcal{C}^+) \prec \mathbf{W}^{complete}\}$,
   The graph $\mathbf{W}(\mathcal{C}^+)$ represents the graph in which was added C edges in a spanning tree of the graph $\mathcal{G}$. An example of this class could be seen in figure ref fig: classe5.

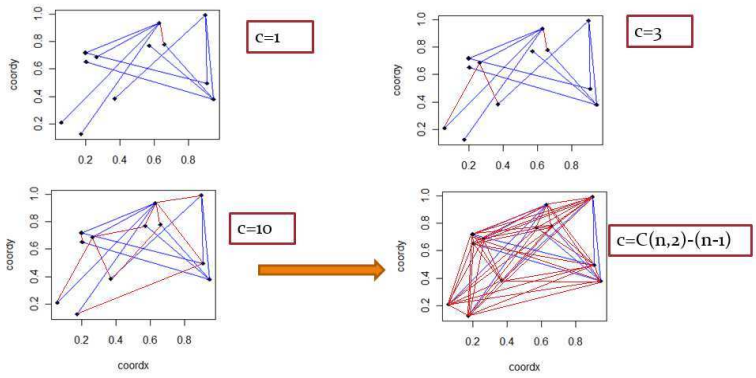


Figure 12: Example of class $\mathcal{W}_5$.

We propose the following distributions to hyper-parameter $\mathcal{C}$:

- $\mathcal{C} \sim U_d\Big(0, \Big[\binom{n}{2} - (n-1)\Big]\Big)$.
- $\mathcal{C} \sim Binomial\Big(\Big[\binom{n}{2} - (n-1)\Big], p\Big)$. The hyper-parameter p allows more flexibility in choosing the number of edges $\mathcal{C}$ to be add from spanning tree.

Despite the classes $\mathcal{W}_2$ and $\mathcal{W}_3$ look the same, there is a very important difference between them. In class $\mathcal{W}_2$, edges are based on boundaries, i.e., if there is any edge between two areas, it means that they are necessarily adjacent. In the class $\mathcal{W}_3$, this is not a rule. There is a chance that both areas are connected by an edge and they do not divide any borders.

4. $\mathcal{W}_4 = \{\mathbf{W}(\mathcal{C}) : \mathbf{W}(\mathcal{C}^-) \prec \mathbf{W}^{adj}\}$, where hyper-parameter $\mathcal{C}$ could be distributed in the following ways:
   - $\mathcal{C} \sim U_d\Big(0, \mathcal{C}_{max}\Big)$, where $\mathcal{C}_{max}$ is the number of edges in adjacency graph,
   - $\mathcal{C} \sim Binomial\Big(\mathcal{C}_{max}, p\Big)$.
5. $\mathcal{W}_5 = \{\mathbf{W}(\mathcal{C}) : \mathbf{W}(\mathcal{C}^-) \prec \mathbf{W}^{completo}\}$, where hyper-parameter $\mathcal{C}$ could be distributed in the following ways:
   - $\mathcal{C} \sim U_d\Big(0, \binom{n}{2}\Big)$,
   - $\mathcal{C} \sim Binomial\Big(\binom{n}{2}, p\Big)$.

   The classes $\mathcal{W}_4$ and $\mathcal{W}_5$ are variations of previous classes. The modification made in these classes allows all edges of the graph to be removed. We can note, however, that when using these classes, the resulting graphs are not always connected.
6. $\mathcal{W}_6 = \{\mathbf{W}(r) : \mathcal{T}_{\mathcal{AG}} \prec \mathbf{W}(r) \prec \mathbf{W}^{complete}\}$, where $\mathbf{W}(r)$ corresponds to a graph in which all vertices are connected to other exactly r vertices. The hyper-parameter r could be:
   - $r \sim U_d\Big(0, \binom{n-1}{2}\Big)$,
   - $r \sim Binomial\Big(\binom{n-1}{2}, p\Big)$.

6.2. Class of random matrices with covariates

In second group we propose some classes that include covariates such as population sizea, area size and gross domestic product (GDP).

This class allows using other important informations to building a neighborhood matrix.

1. $\mathcal{W}_7 = \{\mathbf{W}(h) : W_{ij}(h) = 1 \quad \text{if} \quad g(x_i, x_j) \geq h, \quad W_{ij}(h) = 0 \quad \text{otherwise}\}$, where $x = \{x_1, \ldots, x_n\}$ is the observed vector of covariate X, $g(x_i, x_j) = |x_i - x_j|$ and h is a hyper-parameter.
   Suppose three equidistant towns A, B and C. The B and C cities are small towns with no so good infrastructure, while the city A is a big city. When an individual gets sick in the small towns, he will be brought to the big city. Thus, the problem with the infrastructure provides a of "link" between the small towns and the big city. Thus, city A is neighbor of B and C and the latter two are not neighbors to each other. In this case, covariate X could represents the population of each city. In this way, when the difference between the sizes is greater than a limit h, these cities are linked.

6.3. Combination Class of random matrices

There are situations where it is not sufficient to use a single class. We could analyze, for instance, the delivery traffic (mail) in a region. We assume, in this case, two types of distribution logistics. The first uses road routes, where companies tend to choose shorter routes between cities. This type of distribution can be expressed using the class $\mathcal{W}_8$ of minimum spanning tree, in which costs of edges may be associated with distances between cities (see first proposed class in section 2). The second type of delivery is air traffic between major cities. This type of distribution is essential, since the delivery volume from large cities is far superior when compared to the volume of small cities. Distribution via aircraft can be represented by the class $\mathcal{W}_9$ (see the third proposed class in section 2). Thus, it becomes essential to use both classes, see figure 13 and the alternative for these cases is the combination of classes.

1. $\mathcal{W}_{10} = \{\mathbf{W} : \lambda_1 \mathbf{W}_8 + \lambda_2 \mathbf{W}_9\}$, where $\mathbf{W}_8 \in \mathcal{W}_8$, $\mathbf{W}_9 \in \mathcal{W}_9$, $\lambda_i \in (0, 1)$ with $i = 1, 2$ and $\lambda_1 + \lambda_2 = 1$.
   The parameters λ_i $i = 1, 2$ control the weight that are given to each neighborhood matrix. In our example, the matrix $\mathbf{W}_8$ is connected

and the matrix $\mathbf{W}_9$ is not connected. You can add additional matrix in this combination, the still restrictions are maintained $\sum_i \lambda_i = 1$ and $\lambda_i \in (0, 1)$.

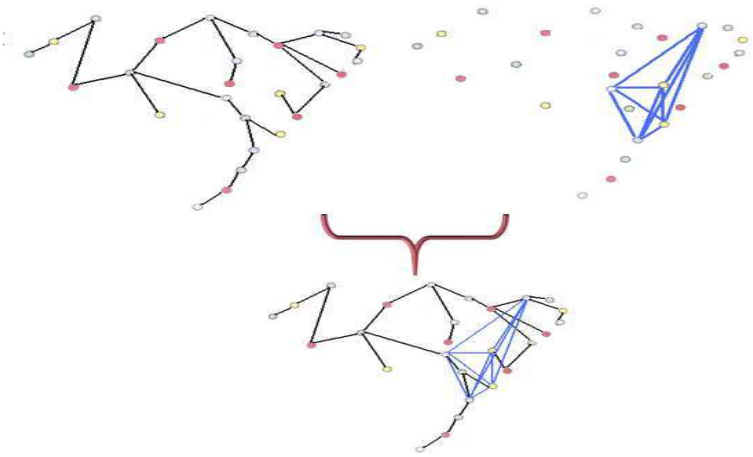


Figure 13: Example of class $\mathcal{W}_{10}$.

References

[1] Walter, S. Disease mapping: a historical perspective. *Spatial epidemiology: methods and applications* 2000; :223–239.

[2] Bithell, J. A classification of disease mapping methods. *Statistics in Medicine* 2000; **19**(17-18):2203–2215. URL `http://onlinelibrary.wiley.com/doi/10.1002/1097-0258(20000915/30)19:17/18%3C2203::AID-SIM564%3E3.0.CO;2-U/full`.

[3] Diggle, PJ. Overview of statistical methods for disease mapping and its relationship to cluster detection. *Spatial Epidemiology: Methods and Applications* 2000; :87–103.

[4] Lawson, AB. Disease map reconstruction. *Statistics in Medicine* 2001; **20**(14):2183–2204. URL `http://onlinelibrary.wiley.com/doi/10.1002/sim.933/full`.

[5] MacNab, YC, Kmetic, A, Gustafson, P, Sheps, S. An innovative application of bayesian disease mapping methods to patient safety research: A canadian adverse medical event study. *Statistics in medicine* 2006;

25(23):3960–3980. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.2507/abstract.

[6] Yu, HL, Chiang, CT, Lin, SD, Chang, TK. Spatiotemporal analysis and mapping of oral cancer risk in changhua county (taiwan): an application of generalized bayesian maximum entropy method. *Annals of epidemiology* 2010; **20**(2):99–107. URL http://www.sciencedirect.com/science/article/pii/S1047279709003421.

[7] MacNab, YC, Gustafson, P. Regression b-spline smoothing in bayesian disease mapping: with an application to patient safety surveillance. *Statistics in medicine* 2007; **26**(24):4455–4474. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.2868/abstract.

[8] Cressie, N, Read, TR. Spatial data analysis of regional counts. *Biometrical Journal* 1989; **31**(6):699–719. URL http://onlinelibrary.wiley.com/doi/10.1002/bimj.4710310607/abstract.

[9] Clayton, D, Kaldor, J. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–681. URL http://www.jstor.org/stable/10.2307/2532003.

[10] Marshall, RJ. Mapping disease and mortality rates using empirical bayes estimators. *Applied Statistics* 1991; :283–294URL http://www.jstor.org/stable/10.2307/2347593.

[11] Tsutakawa, RK, Shoop, GL, Marienfeld, CJ. Empirical bayes estimation of cancer mortality rates. *Statistics in Medicine* 1985; **4**(2):201–212. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.4780040210/abstract.

[12] Mollie, A, Richardson, S. Empirical bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine* 1991; **10**(1):95–112. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.4780100114/abstract.

[13] Besag, J, York, J, Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; **43**(1):1–20. URL http://link.springer.com/article/10.1007/BF00116466.

[14] Carlin, BP, Banerjee, S. Hierarchical multivariate car models for spatio-temporally correlated survival data. *Bayesian statistics* 2003; **7**:45–63.

[15] Jin, X, Carlin, BP. Multivariate parametric spatiotemporal models for county level breast cancer survival data. *Lifetime Data Analysis* 2005; **11**(1):5–27. URL http://link.springer.com/article/10.1007/s10985-004-5637-1.

[16] Gelfand, AE, Vounatsou, P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 2003; **4**(1):11–15. URL http://biostatistics.oxfordjournals.org/content/4/1/11.short.

[17] Held, L, Natário, I, Fenton, SE, Rue, H, Becker, N. Towards joint disease mapping. *Statistical methods in medical research* 2005; **14**(1):61–82. URL http://smm.sagepub.com/content/14/1/61.short.

[18] Held, L, Graziano, G, Frank, C, Rue, H. Joint spatial analysis of gastrointestinal infectious diseases. *Statistical methods in medical research* 2006; **15**(5):465–480. URL http://smm.sagepub.com/content/15/5/465.short.

[19] Martínez-Beneito, M, López-Quilez, A, Botella-Rocamora, P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in medicine* 2008; **27**(15):2874–2889. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.3103/abstract.

[20] Knorr-Held, L, Best, NG. A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2001; **164**(1):73–85. URL http://onlinelibrary.wiley.com/doi/10.1111/1467-985X.00187/abstract.

[21] Sun, D, Tsutakawa, RK, Kim, H, He, Z, et al. Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 2000; **19**(15):2015–2035.

[22] Silva, GL, Dean, C, Niyonsenga, T, Vanasse, A. Hierarchical bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in medicine* 2008; **27**(13):2381–2401. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.3094/abstract.

[23] Kafadar, K, Freedman, LS, Goodall, CR, Tukey, JW. Urbanicity-related trends in lung cancer mortality in us counties: white females and white males, 1970–1987. *International journal of epidemiology* 1996; **25**(5):918–932. URL http://ije.oxfordjournals.org/content/25/5/918.short.

[24] Goodall, CR, Kafadar, K, Tukey, JW. Competing and using moral versus urban measures in statistical applications. *The American Statistician* 1998; **52**(2):101–111. URL http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1998.10480548.

[25] Kafadar, K. Geographic trends in prostate cancer mortality: an application of spatial smoothers and the need for adjustment. *Annals of epidemiology* 1997; **7**(1):35–45. URL http://www.sciencedirect.com/science/article/pii/S1047279796001019.

[26] Jackson, MC, Huang, L, Xie, Q, Tiwari, RC. A modified version of moran's i. *International Journal of Health Geographics* 2010; **9**:33.

[27] White, G, Ghosh, SK. A stochastic neighborhood conditional autoregressive model for spatial data. *Computational statistics & data analysis* 2009; **53**(8):3033–3046. URL http://www.sciencedirect.com/science/article/pii/S0167947308003885.

[28] Mollié, A. Bayesian mapping of disease. *Markov chain Monte Carlo in practice* 1996; **1**:359–379.

[29] Brook, D. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika* 1964; **51**(3/4):481–483. URL http://www.jstor.org/stable/10.2307/2334154.

[30] Kirchhoff, G. Über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. *Ann. Phys. Chem* 1847; **72**:497–508.

[31] Khan, K, Arino, J, Hu, W, Raposo, P, Sears, J, Calderon, F, Heidebrecht, C, Macdonald, M, Liauw, J, Chan, A, et al. Spread of a novel influenza a (h1n1) virus via global airline transportation. *New England Journal of Medicine* 2009; **361**(2):212–214. URL http://www.nejm.org/doi/full/10.1056/NEJMc0904559.

[32] Warren, A, Bell, M, Budd, L. Airports, localities and disease: representations of global travel during the h1n1 pandemic. *Health & place* 2010; **16**(4):727–735. URL http://www.sciencedirect.com/science/article/pii/S1353829210000328.

[33] Hsu, CI, Shih, HH. Transmission and control of an emerging influenza pandemic in a small-world airline network. *Accident Analysis & Prevention* 2010; **42**(1):93–100. URL http://www.sciencedirect.com/science/article/pii/S0001457509001717.