

Bayesian analysis of spatial data using different variance and neighbourhood structures

Renato Couto Rampaso, Aparecida Doniseti Pires de Souza & Edilson Ferreira Flores

To cite this article: Renato Couto Rampaso, Aparecida Doniseti Pires de Souza & Edilson Ferreira Flores (2016) Bayesian analysis of spatial data using different variance and neighbourhood structures, Journal of Statistical Computation and Simulation, 86:3, 535-552, DOI: 10.1080/00949655.2015.1022549

To link to this article: <https://doi.org/10.1080/00949655.2015.1022549>



Published online: 16 Mar 2015.



Submit your article to this journal [↗](#)






Article views: 322



View Crossmark data [↗](#)

Bayesian analysis of spatial data using different variance and neighbourhood structures

Renato Couto Rampaso* , Aparecida Doniseti Pires de Souza  and
Edilson Ferreira Flores 

Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, SP, Brazil

(Received 13 August 2014; accepted 19 February 2015)

In disease mapping, the overall goal is to study the incidence or mortality risk caused by a specific disease in a number of geographical regions. It is common to assume that the response variable follows a Poisson distribution, whose average rate can be explained by a group of covariates and a random effect. For this random effect, it is considered conditional autoregressive (CAR) models, which carry information about the neighbourhood relationship between the regions. The focus of this paper was to explore and compare some CAR models proposed in the literature. An application with epidemiological data was conducted to model the risk of death due to Crohn's Disease and Ulcerative Colitis in the State of São Paulo – Brazil. Finally, a simulation study was done to strengthen the results and assess the performance of the models in the presence of various levels of spatial dependence.

Keywords: conditional autoregressive models; disease mapping; spatial Bayesian inference

AMS Subject Classifications: 62H11; 62F15

1. Introduction

Understanding the spatial distribution of data from random phenomena is a great challenge to clarify important issues in various areas of knowledge. From the point of view of statistics, the interest is to develop probabilistic models that adequately represent the spatial distribution of the phenomenon under study. Application examples involve the mapping of diseases or crime in a region of interest, modelling of air pollutants in an urban centre, satellite images analysis, field trials and so forth.

In disease mapping, the overall goal is to study the incidence or mortality risk caused by a particular disease in a number of geographical regions. It has great significance in epidemiological studies and the definitions of public policies, identifying areas presenting high rates and possible risk factors associated with it. Stern and Cressie [1] stated that maps which have simple ratios, such as the ratio between the number of deaths and the population in the study area, have been criticized for not being reliable due to the non-constant variance associated with heterogeneity in population sizes.

An alternative to areal data analysis is taking models that smooth the relative risk. Since the random variable of interest is generally a count that represents the occurrence of a given random

*Corresponding author. Email: renato.rampaso@tjdf.tjus.br

phenomenon, it is common to assume that the response variable follows a Poisson distribution, which the average rate can be explained by a group of covariates and a random effect. For this random effect, it is considered a prior distribution that carries information about the neighbourhood relationship between the areas of a region of interest.

A popular class of models used to represent the spatial random effects is the conditional autoregressive (CAR). Specification of these models is directly linked to its variance matrix. This matrix is extremely important because it allows the incorporation of spatial structure in the model. Different specifications for this matrix result in different CAR models. A necessary imposition to define a valid model is that the matrix is symmetric and positive-definite.

In this context, Lee [2] presented a study involving four different CAR models: intrinsic (ICAR), convolution, Cressie and Leroux. The objective of this paper is to compare these CAR models through an application and a simulation study. Furthermore, we complement with other approaches to the representation of the spatial neighbourhood structure, which allows estimating components of the adjacency matrix. The scarcity of studies that compare different CAR models and their variance structures for the random effects justifies performing this work. Here, we listed the characteristics of each model, the advantages and disadvantages that will enable the researcher to acquire enough knowledge to choose the most appropriate that meets their needs.

In Section 2, the models are presented, with their definitions, formulations, characteristics and peculiarities. Inference procedures are described in Section 2.3. In Section 3, an application with real data using CAR models is conducted. It refers to the number of deaths caused by Crohn's Disease and Ulcerative Colitis in São Paulo in the period between 2008 and 2012. To strengthen the results and in order to ascertain the performance of the models in the presence of various levels of spatial dependence, a simulation study is presented in Section 4. Finally, in Section 5, an overview of the results and final considerations are raised.

2. CAR models

Before presenting the general formulation of the class of Bayesian hierarchical models for disease mapping and the CAR models proposed in the literature, we will draw a brief parallel to other existing proposals to handle spatial data.

Basically, there are two different ways to specify a spatially structured model. One is the geostatistical approach to spatial process. In this approach, it is common to assume that the data are related to points, observed in the centre or centroid of each region and the distances between these points are used to represent the spatial structure through a covariance function. Such covariance functions must follow some mathematical conditions for admitting permissible forms (see [3,4] for further study on the permissibility of covariance functions). Haining [5] cites some examples of covariance functions, including the triangular correlation function, the spherical model and the exponential function. Wall [6] cites a common problem in using the geostatistical approach, which is the arbitrariness of assigning the summary for the whole region to the centroid. On the other hand, an advantage of this modelling is that the spatial covariance function is modelled directly, and its structure has a simple and direct understanding.

The other way to specify a spatial model takes into account a structure of discrete index neighbourhood. Here, each observation is not tied to only one point, but to an entire region. In this approach, instead of defining a centroid, we use a matrix that tells us what the neighbours of each region are. Once such neighbourhood structure is defined, autoregressive models (similar to discrete temporal series models) are adopted to represent the spatial random effects. One of the most popular is the class of CAR models. These models have proved very useful in the context of Bayesian inference, used as prior distributions for random effects in hierarchical models.

In subsequent subsections, an overview about disease mapping and different CAR models from literature will be defined.

2.1. Disease mapping and the Bayesian approach

In disease mapping, the interest is to investigate the incidence or risk of certain disease in n continuous areas. Such risks are often arranged in maps that help visualize the distribution of the incidence, detect areas with high risks and possible evidence of spatial association. For example, one may be interested to see the number of deaths or incidence of certain types of cancer over the regions of a state.

According to Alan E Gelfand and Montserrat Fuentes [7] disease mapping refers to a collection of methods extending to the small areas estimation to directly utilize the spatial setting and assumed positive spatial correlation between observations, essentially borrowing more information from neighbouring areas than from areas far away and smoothing local rates towards local, neighbouring values. The term disease mapping first appeared in [8] which defined empirical Bayesian methods constructed from a Poisson regression with random intercepts defined with spatial correlation.

For the Poisson regression models, let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ be the vector of observed values of a particular disease. It is known that the occurrence of certain diseases is linked to known risk factors, such as age and sex distribution of the population. Thus, to remove the effect that these differences may have on the incidence, the expected values, denoted by $\mathbf{E} = (E_1, E_2, \dots, E_n)$, are calculated, considering the demographic structure site. The vector \mathbf{E} can be seen as expected values under the hypothesis of constant risk between regions.

The simplest measure to assess the risk of a disease is through the standardized mortality ratio (SMR), calculated for a given area i by $SMR_i = Y_i/E_i$, $i = 1, 2, \dots, n$. Values greater than one imply a risk of disease higher than expected, while below one indicate a risk lower than expected for the respective area. However, a low value of E_i can happen if the population of a certain place is too low or if the disease under study is rare, which implies a non-plausible high risk for the region in question.

To overcome this problem, Bayesian spatial models can be adopted. Such models allow the use of covariates that can provide information on the risk of mortality, as well as a set of random effects that capture the dependence between neighbouring regions. A general formulation of this class of models is given by

$$\begin{aligned} Y_i | E_i, R_i &\sim \text{Poisson}(E_i R_i) \\ \log(R_i) &= \alpha + x_i^t \beta + \phi_i, \quad i = 1, 2, \dots, n. \end{aligned} \quad (1)$$

where R_i denotes the risk of disease for area i , which is estimated considering an intercept α common to all regions, a set of p covariates $x_i^t = (x_{i1}, \dots, x_{ip})$ and a random effect ϕ_i . The prior distribution used for the parameter α is a *flat*, which corresponds to an improper uniform $U(-\infty, +\infty)$ throughout the entire real line. More details justifying the choice of this prior are provided in Section 2.2.1. Parameters of the coefficient vector β are generally assumed by normal prior distributions with low precision.

In Poisson distribution, $\text{Var}(Y_i) = E(Y_i)$, although in practical cases $\text{Var}(Y_i) > E(Y_i)$ occur. This scenario is known as overdispersion. The random effect ϕ_i seeks to capture the spatial dependence or some overdispersion present in the data. An overdispersion can occur due to the existence of not measured covariates or risk factors.

2.2. CAR model specification

Since it is flexible and easy to handle, the multivariate normal distribution is usually used to represent the joint distribution of the random effects. It may be assumed then,

$$\phi \sim \text{NM}(\mu, \Sigma(\theta)), \tag{2}$$

where μ is the mean vector and $\Sigma(\theta)$ is the variance matrix which considers the spatial structure through a set of parameters θ .

Let n be the total number of areas in study. In general, it is considered that

$$\Sigma(\theta) = \sigma^2 \Phi; \quad \Phi = (\mathbf{I} - \rho \mathbf{C})^{-1} \mathbf{M}, \tag{3}$$

where σ^2 is a general variance parameter, \mathbf{I} is an identity matrix $n \times n$, ρ is a parameter that measures the spatial dependence, $\mathbf{C} = C_{ij}$ is a spatial association matrix with zeros on its diagonal, and $\mathbf{M} = M_{ii}$ is a diagonal matrix.

Models for ϕ in Equation (2) are usually specified by n full conditional distributions $p(\phi_i | \phi_{-i})$, where $\phi_{-i} = (\phi_1, \dots, \phi_{i-1}, \phi_{i+1}, \dots, \phi_n)$. To set a valid model, one should specify C_{ij} and M_{ii} such that the matrix $\Phi = (\mathbf{I} - \rho \mathbf{C})^{-1} \mathbf{M}$ is symmetric and positive-definite.

Inspecting Φ^{-1} , note that Φ is symmetric if $m_{ij}c_{ij} = m_{ii}c_{ji}$. Moreover, it is easy to see that $\Phi = \mathbf{M}^{1/2}(\mathbf{I} - \rho \mathbf{M}^{-1/2} \mathbf{C} \mathbf{M}^{1/2})^{-1} \mathbf{M}^{1/2}$. Therefore, the matrix of variances is positive-definite for $\rho \in (\rho_{\min}, \rho_{\max})$, where $1/\rho_{\min}$ e $1/\rho_{\max}$ are the smallest and largest eigenvalue of the matrix $\mathbf{M}^{-1/2} \mathbf{C} \mathbf{M}^{1/2}$, respectively. In practice, we expect a positive spatial dependence. Thus, it is possible to limit the range of ρ at $(0, \rho_{\max})$. A value of $\rho = 0$ implies spatial independence.

Let $j \sim i$ denoting the presence of neighbourhood between areas j and i . Two areas are said neighbours if they share common border, although other ways of defining neighbourhoods can be adopted (see, e.g. [9, p.384–385]). The full conditional distributions of a CAR model can be generally defined as

$$\phi_i | \phi_{-i} \sim N \left(\mu_i + \rho \sum_{j \sim i} c_{ij} (\phi_j - \mu_j), \sigma^2 m_{ii} \right). \tag{4}$$

2.2.1. Intrinsic model

One of the simpler CAR models is the intrinsic (ICAR, shortening), proposed by Besag et al. [10] where the variance matrix Φ is not positive-definite. For this reason, the model is improper and can only be used as prior distribution rather than likelihood to the data. This model is obtained from Equation (4) doing $c_{ij} = 1/n_i$ if areas i and j are adjacent and 0 otherwise, $m_{ii} = 1/n_i$, n_i the number of neighbours of area i and $\rho = \rho_{\max}$, which under these specifications of c_{ij} and m_{ii} implies $\rho = 1$. Here, matrix \mathbf{C} is equivalent to the classical normalized weights matrix. Without loss of generality, it is also assumed $\mu_i = 0, i = 1, 2, \dots, n$. Its full conditional distribution can be expressed as

$$\phi_i | \phi_{-i} \sim N \left(\frac{1}{n_i} \sum_{j \sim i} \phi_j, \frac{\sigma^2}{n_i} \right). \tag{5}$$

The conditional expectation of random effect ϕ_i is the average of the effects of its neighbours. Its variance structure induces that the higher the number of neighbours of area i , n_i , the greater the precision of its conditional variance.

Because it is an improper model, some caution is needed. In general, it is convenient to assume that the effects distribution have zero mean. In addition, a constraint is necessary for the

random effects to sum zero ($\sum_{i=1}^n \phi_i = 0$). Normally, this constraint is done numerically recentering samples ϕ_i around their own mean.[11] Besag and Kooperberg [12] showed that constrain the effects to sum to zero and specify a separate intercept with the invariant prior distribution $U(-\infty, +\infty)$ is equivalent to the model without constraint and without intercept. Therefore, using a model with intercept requires a flat prior to it.[13] Imposed such restrictions, the joint distribution of the random vector ϕ is multivariate normal with mean vector 0 and variance matrix (singular) $\sigma^2 \mathbf{D}^-$ (where \mathbf{D}^- is the generalized inverse of \mathbf{D}), with the ij th element of \mathbf{D} defined as

$$d_{ij} = \begin{cases} n_i & \text{se } j = i, \\ -1 & \text{se } j \sim i, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The ICAR model considered has some disadvantages. The strength of spatial dependence between random effects is always considered maximum ($\rho = 1$). Furthermore, the variance parameter σ^2 is used both to capture overdispersion and spatial dependence.[14]

2.2.2. Convolution model

Another popular model for random effects is the convolution model. Also proposed by Besag et al.,[10] it combines the intrinsic model (spatially structured) with an additional random effect (spatially unstructured). This model is given by

$$\begin{aligned} \phi_i &= \theta_i + \psi_i, \\ \theta_i &\sim N(0, \sigma_\theta^2), \\ \psi &= (\psi_1, \dots, \psi_n) \sim \text{ICAR}(W, \sigma_\psi^2). \end{aligned} \quad (7)$$

The term ψ has prior distribution ICAR described in Equation (5). An explanation for the inclusion of θ_i is taken from the empirical study of Breslow[15] who observed an additional extra variation in log-linear Poisson model. That is, the random effect $\theta = (\theta_1, \dots, \theta_n)$ – independent between areas – is added in order to absorb this additional variation not captured by the intrinsic spatial random effect. Thus, by varying the relative size of these two random components, it is expected to capture different levels of spatial correlation. The name convolution is because the density of the random effects ϕ_i 's will be the convolution of joint probability densities of the vectors θ and ψ . [16]

Some criticisms are pointed to the convolution model. Each observation is represented by two random effects, but only the sum $\theta_i + \psi_i$ is identifiable. Furthermore, Eberly et al. [17] points some convergence problems in estimation, which can be affected by the choice of prior distributions for the parameters and even the chain chosen for the initial values.

2.2.3. Cressie model

The Cressie model, also known as proper CAR, is an alternative to capture different levels of spatial dependence. Their initial ideas can be found in [9], with additional details in [1]. The idea here is to introduce ρ as a parameter to be estimated.

For matrices \mathbf{C} and \mathbf{M} , will be considered here the same specification imposed in ICAR model: $c_{ij} = 1/n_i$ if areas i and j are adjacent and 0 otherwise, and $m_{ii} = 1/n_i$. With these settings, the upper limit for ρ is 1. Since the trace of matrix $\mathbf{M}^{-(1/2)} \mathbf{C} \mathbf{M}^{1/2}$ equals zero, then $\rho_{\min} < 0 < \rho_{\max}$ (since $\mathbf{M}^{-(1/2)} \mathbf{C} \mathbf{M}^{1/2}$ is a square matrix with zeros on its diagonal and $\text{tr}(\mathbf{M}^{-(1/2)} \mathbf{C} \mathbf{M}^{1/2}) = \sum_{i=1}^n \delta_i = 0$, δ_i the i th eigenvalue of matrix $\mathbf{M}^{-(1/2)} \mathbf{C} \mathbf{M}^{1/2}$). Therefore, a plausible range that

makes this a proper model is $0 \leq \rho < 1$, $\rho = 1$ implying the ICAR model. Thus, the set of random effects $\phi = (\phi_1, \dots, \phi_n)$ has a multivariate normal distribution with precision matrix $\sigma^2 \mathbf{Q}^{-1}$, that is,

$$\phi \sim NM(\mu, \sigma^2 \mathbf{Q}^{-1}), \quad 0 \leq \rho < 1, \tag{8}$$

where ij th element of matrix \mathbf{Q} is defined as

$$q_{ij} = \begin{cases} n_i & \text{se } j = i, \\ -\rho & \text{se } j \sim i, \\ 0 & \text{otherwise.} \end{cases}$$

Univariate full conditional distribution for the effects ϕ_i is then

$$\phi_i | \phi_{-i} \sim N \left(\rho \frac{1}{n_i} \sum_{j \sim i} \phi_j, \frac{\sigma^2}{n_i} \right). \tag{9}$$

The conditional variance is the same as intrinsic model. The conditional expectation of this model can be seen as a weighted average of local random effects mean (with weight ρ) and zero overall average (with weight $1 - \rho$)

$$E(\phi_i | \phi_{-i}) = (1 - \rho) \times 0 + \rho \times \frac{1}{n_i} \sum_{j \sim i} \phi_j.$$

For $\rho = 0$ we have independence, since the values will remain around a zero mean. Moreover, values of ρ near one indicate strong spatial correlation.

It is noteworthy, however, that there is some flexibility in the specification of the matrices \mathbf{C} and \mathbf{M} . Stern and Cressie,[1] for example, considered $C_{ij} = \sqrt{E_j/E_i}$ if areas i and j are adjacent and 0 otherwise, and $M_{ii} = 1/E_i$. This specification, however, changes the allowable variation of ρ and we have to get a new range to achieve a proper model.

A drawback with this model is due to the fact that in the absence of spatial dependence between random effects (ρ close to zero), the conditional variance is invariant and continues depending on the number of neighbours n_i .

2.2.4. Leroux model

Another model, more general than Cressie, was proposed by Leroux et al.[14] The set of effects ϕ is represented by a multivariate normal distribution:

$$\phi \sim NM(\mathbf{0}, \sigma^2[\rho \mathbf{D} + (1 - \rho)\mathbf{I}]^{-1}). \tag{10}$$

The precision matrix is $\mathbf{L} = \rho \mathbf{D} + (1 - \rho)\mathbf{I}$, where \mathbf{I} denotes an identity matrix of order n and the matrix \mathbf{D} is the same as already defined in Equation (6). It is easy to see that for $\rho = 0$, $\mathbf{L} = \mathbf{I}$ and we have a model with independent random effects. On the other hand, when $\rho = 1$, $\mathbf{L} = \mathbf{D}$ and we obtain the intrinsic model. If $0 \leq \rho < 1$, the joint distribution (10) is proper. The univariate full conditional distribution is then given by

$$\phi_i | \phi_{-i} \sim N \left(\frac{\rho}{n_i \rho + 1 - \rho} \sum_{j \sim i} \phi_j, \frac{\sigma^2}{n_i \rho + 1 - \rho} \right). \tag{11}$$

The conditional variance has now a more attractive form than the model (9). For ρ close to 1, the conditional variance is close to σ^2/n_i . When $\rho = 0$, the conditional variance comes down to σ^2 ,

no longer depending on the number of neighbours n_i . The conditional expectation can be seen as a weighted average of the local mean of the random effects (with weight $n_i\rho$) and the overall average 0 (with weight $1 - \rho$), that is,

$$E(\phi_i|\phi_{-i}) = \frac{1 - \rho}{n_i\rho + 1 - \rho} \times 0 + \frac{n_i\rho}{n_i\rho + 1 - \rho} \times \frac{1}{n_i} \sum_{j \sim i} \phi_j.$$

Furthermore, the conditional variance can be also seen as a weighted average of the local variance of the intrinsic model (with weight $1 - \rho$) and the variance of an independent model (with weight $n_i\rho$)

$$V(\phi_i|\phi_{-i}) = \frac{1 - \rho}{n_i\rho + 1 - \rho} \times \sigma^2 + \frac{n_i\rho}{n_i\rho + 1 - \rho} \times \frac{\sigma^2}{n_i}.$$

Rodrigues [18] and Banerjee et al. [19] pointed out that proper CAR models can induce little correlation between neighbouring areas. This disadvantage can be one of the reasons that makes the intrinsic and convolution models most popular choices to accommodate the spatial random effects.

2.2.5. Lu model

Lu et al. [20] developed a model in which the adjacency matrix is no longer deterministic, but random. In other words, the variance structure of the model is now estimated. This is done by modelling the weights of the adjacency matrix as

$$w_{ij}|p_{ij} \sim \text{Bernoulli}(p_{ij}), \tag{12}$$

where

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \mathbf{z}'_{ij}\gamma.$$

Here, \mathbf{z}_{ij} is a set of known inherent characteristics of regions i and j , with corresponding vector parameter γ . Such characteristics may include, for example, the distance between the centroids of each region. Now, the areas i and j are considered neighbours with probability p_{ij} , as long as they share common border. The Lu model allows the neighbourhood structure to be determined by the value of the process in each area and variables that measure how similar two regions are. Taking the simplest case – and used in this paper – with a presence of only one covariate, the model is then defined as

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \gamma_0 + \gamma_1 z_{ij}. \tag{13}$$

Models of this nature are especially useful in spatial boundary analysis. In this type of analysis, we determine boundaries on a map that separate areas alternating high and low values. This method is often called *wombling*, referring to a pioneer paper of Womble.[21]

The specification for the random vector effects refers to Equations (2) and (3) and the intrinsic model, that is,

$$\phi \sim NM(0, \sigma^2(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}),$$

where $c_{ij} = w_{ij}/\sum_j w_{ij}$ to i and j adjacent and 0 otherwise, and $m_{ii} = 1/\sum_j w_{ij}$, \mathbf{W} the adjacency matrix. One problem is that the determinant of the precision matrix $\mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})/\sigma^2 = (\mathbf{M} - \mathbf{W})/\sigma^2$ is zero for cases where an area has no neighbours. In this model, such a scenario can happen even in the case where there are no ‘islands’ in the studied map, since the random nature

for the adjacency matrix allows $w_{ij} = 0$ even though i and j share common border. To overcome this problem, Lu et al. [20] proposed an approach that results in an appropriate estimation for the determinant of the precision matrix. This is done replacing \mathbf{M} by \mathbf{M}^* , adding a small positive value ϵ to the diagonal elements of \mathbf{M} . Thus, the matrix $(\mathbf{M}^* - \mathbf{W})/\sigma^2$ is diagonal dominant, symmetric and positive-definite.

A drawback pointed to this model is that the parameter vector γ – effects of the covariates in the logistic regression model – not always well identifies the data. In some cases, prior distributions with informative content are necessary to ensure the desired effect by covariates \mathbf{z}_{ij} in the boundary analysis.

2.3. Model specification and inference procedures

Considering the models proposed in Equations (1) and (2), it is also necessary to specify the prior distributions for ρ (Cressie and the Leroux models), γ (Lu model) and σ (all models). Following the work of Gelman,[22] to the standard deviation was specified a non-informative prior with uniform distribution, $p(\sigma) \sim U(0, T)$. Lee [2] performed some sensitivity studies to the value of T , pointing out that, since that is a sufficiently large value, this choice does not alter the estimates produced by the posterior distributions. So, we chose to set $T = 10$.

For parameter ρ (Cressie and Leroux models), we previously defined the interval (0,1), such that the models are proper. Therefore, a natural choice of prior distribution is $p(\rho) \sim U(0, 1)$.

Lu et al. [20] suggests using normal prior distributions for γ_0 and γ_1 , specifying

$$p(\gamma_0) \sim N(\mu_0, \sigma_0^2) \quad \text{and} \quad p(\gamma_1) \sim N(\mu_1, \sigma_1^2).$$

The parameter ϵ , used to make the determinant of the variance matrix in Lu model positive-definite, was fixed at 0.5. Lu et al. [20] shows that values of ϵ approaching zero increase bias in risks estimation. On the other hand, values close to one increase σ^2 bias. So the above choice for ϵ represents an intermediate point to balance the faced problem.

Assuming the model specified in Equations (1) and (2) with $\lambda_i = E_i R_i$, $i = 1, 2, \dots, n$, the likelihood function is given by

$$\begin{aligned} L(\alpha, \beta, \phi; y) &= \prod_{i=1}^n \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} = \prod_{i=1}^n \frac{(E_i R_i)^{y_i} \exp(-E_i R_i)}{y_i!} \\ &= \prod_{i=1}^n \frac{[E_i \exp(\alpha + x_i^t \beta + \phi_i)]^{y_i} \exp(-E_i \exp(\alpha + x_i^t \beta + \phi_i))}{y_i!}. \end{aligned} \tag{14}$$

Therefore, the joint posterior distributions for the parameter vector of the models, assuming independence between prior distributions, are as follows :

(i) For intrinsic model:

$$\begin{aligned} p(\alpha, \beta, \phi, \sigma | y) &\propto L(\alpha, \beta, \phi; y) p(\alpha, \beta, \phi, \sigma) \\ &\propto L(\alpha, \beta, \phi; y) p(\alpha) p(\beta) p(\phi) p(\sigma). \end{aligned} \tag{15}$$

(ii) For convolution model:

$$\begin{aligned} p(\alpha, \beta, \phi, \sigma_\psi, \sigma_\theta | y) &\propto L(\alpha, \beta, \phi; y) p(\alpha, \beta, \phi, \sigma_\psi, \sigma_\theta) \\ &\propto L(\alpha, \beta, \phi; y) p(\alpha) p(\beta) p(\phi) p(\sigma_\psi) p(\sigma_\theta). \end{aligned} \tag{16}$$

(iii) For Cressie and Leroux models:

$$\begin{aligned} p(\alpha, \beta, \phi, \sigma, \rho|y) &\propto L(\alpha, \beta, \phi; y)p(\alpha, \beta, \phi, \sigma, \rho) \\ &\propto L(\alpha, \beta, \phi; y)p(\alpha)p(\beta)p(\phi)p(\sigma)p(\rho). \end{aligned} \quad (17)$$

(iv) For Lu model:

$$\begin{aligned} p(\alpha, \beta, \phi, \sigma, W, \gamma|y) &\propto L(\alpha, \beta, \phi; y)p(\alpha, \beta, \phi, \sigma, W, \gamma) \\ &\propto L(\alpha, \beta, \phi; y)p(\alpha)p(\beta)p(\phi)p(\sigma)p(W)p(\gamma). \end{aligned} \quad (18)$$

From Equations (15)–(18), it could be noticed that the expressions for the posterior distributions are complex and do not have closed forms. Thus, to obtain the marginal posterior distributions is necessary to use numerical methods and the Gibbs sampler were used. The estimation was done using OpenBUGS [23] software. The interface used within OpenBUGS was GeoBUGS, an add-on module for modelling spatial data.

Finally, model choice arises in selecting among the five CAR models. As in Gosh et al., [24] we also consider three popular approaches to the models selection, namely the conditional predictive ordinate (CPO, [25]), the expected predictive deviance (EPD, [26]) and the deviance information criterion (DIC). [27] As a complement to evaluate the performance of the models, we also provide the residual sum of squares (RSS). Theoretical details and implementation ways about these criteria are provided in these papers and are not presented here. Rather, for each model, we supply the values for the different measures and present a brief discussion.

3. Epidemiological data application

In this section, an application was explored using models for disease mapping described in Section 2. The collected data are available on Datasus portal (<http://www.datasus.gov.br>). The variable of interest refers to epidemiological data in the state of São Paulo – Brazil, in the period from 2008 to 2012, for the mortality due to Crohn’s Disease and Ulcerative Colitis. The cartographic base considered in this work consists in the administrative division of the state of São Paulo by microregions, according to the Brazilian Institute of Geography and Statistics (IBGE). That choice lies in the fact that many cities in the state have a population smaller than 10,000 inhabitants, and eventually do not present death cases for the given period. Thus, the 645 cities were grouped into 63 microregions. We opted for death by residence, that is, we considered the place where the person lives instead of where his or her death was recorded. The interest in the choice of Crohn’s Disease and Ulcerative Colitis is due to the lack of studies in this context, which calls attention to a better understanding of the spatial distribution of the disease.

Regarding the Lu model, we adopted the proposal made in Equation (13) to estimate the probabilities of two regions being neighbours with one explanatory variable. As a covariate z_{ij} , we have chosen the distance between the centroids of the i and j microregions. Notice that this variable is a measure of dissimilarity. The distance matrix was obtained from a georeferenced database of the regions of the State of São Paulo provided by IBGE. Distances were calculated from tools available in ArcView and ArcGIS packages. Values in the distance matrix were normalized to the interval (0,1).

The hiperpriors proposed for parameters γ_0 and γ_1 based on [20] were $p(\gamma_0) \sim N(1, 1)$ and $p(\gamma_1) \sim N(-5, 1)$.

Although non-informative prior distributions are often used in disease mapping, the use of informative prior distributions was necessary to induce the desired effect in the covariate under study. The negative value of the average γ_1 is used due to the fact that the distance between centroids is a measure of dissimilarity.

3.1. Crohn's Disease and Ulcerative Colitis

Crohn's Disease and Ulcerative Colitis are inflammatory bowel diseases that have certain clinical similarities. Although there is no scientific consensus, both have characteristics resembling an autoimmune disease. In an autoimmune disease, the immune system will produce antibodies that cannot differentiate antigens (invaders such as bacteria and viruses) of healthy tissues. That is, the body attacks itself. Crohn's disease affects mainly the lower part of the small intestine (ileum) and large intestine (colon), but can affect any part of the gastrointestinal tract, from the mouth to the anus. Ulcerative Colitis attacks only the colon and rectum, with inflammation and ulceration of the most superficial layer of the colon.

Causes for the incidence of these diseases are unclear. Among experts it is believed that genetic and environmental factors exert their influence. It is also known that inflammatory bowel diseases are more prevalent in the northern hemisphere and also affect men and women equally. Although they are not infectious diseases, a spatial dependency between the regions can occur due to the presence of environmental factors that are correlated among themselves. As an exploratory analysis, Figure 1 stretches the standardized mortality ratios related to Crohn's Disease and Ulcerative Colitis, calculated for each microregion. From this map, it is already possible to get an overview of how risk behaves over the microregions. However, the map itself does not provide relevant information, such as neighbourhood relations between regions or measurement of the spatial structure, which is why the CAR models are needed.

3.2. Results

Applying the models defined in Equation (1) and Section 2.2, the estimation was carried out via Markov Chain Monte Carlo (MCMC), with 5000 initial iterations discarded for burn-in and a total of 20,000 subsequent iterations. The thin used was equal to 30. Firstly, a study to verify the indication of convergence of the parameters was conducted. Convergence Analysis Diagnosis and Output (CODA) package, which is a set of functions implemented in the R statistical programming language [28] was used for obtaining convergence diagnostic. Moreover, the chain trajectories and autocorrelations graphs were checked. All the computation was developed in an Intel Core i7-2600 3.4 GHz computer, with 8GB RAM memory, which processed about 1369, 917, 1301, 911, 315 iterations per second in the Intrinsic, Convolution, Cressie, Leroux and Lu models, respectively.

Table 1 provides the estimates (posterior means) of each parameter. Note that the intercepts of all models can be considered non-significant, since their credibility intervals contain the zero value. Among the variance parameters (except for σ_θ^2 , representing the variability of

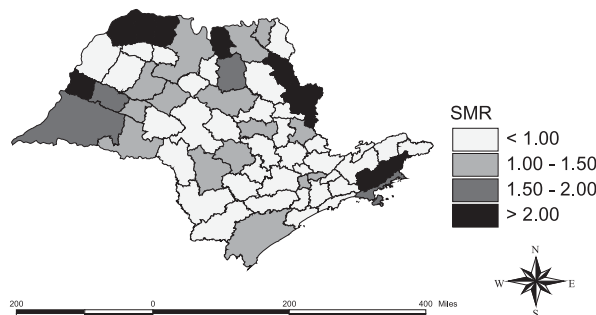


Figure 1. Standardized mortality ratios for Crohn's Disease and Ulcerative Colitis calculated for each microregion of São Paulo.

Table 1. Parameter estimates, standard deviations and 95% credible intervals from data relating to death by Crohn's Disease and Ulcerative Colitis.

Parameter	Mean	Standard deviation	Cr.I.(95%)
α_{In}	0.0044	0.0254	(-0.0460, 0.0054)
α_{Cv}	0.0026	0.0399	(-0.0808, 0.0806)
α_{Cr}	-0.0344	0.1654	(-0.3637, 0.2998)
α_{Le}	0.0058	0.2697	(-0.5305, 0.5157)
α_{Lu}	0.0029	0.1408	(-0.2786, 0.2731)
σ_{In}^2	1.3559	0.2835	(0.9014, 2.0060)
σ_{β}^2	0.0725	0.0748	(0.0012, 0.2807)
σ_{ψ}^2	1.1222	0.3649	(0.3843, 1.8820)
σ_{Cr}^2	1.4440	0.3090	(0.9516, 2.1640)
σ_{Le}^2	1.1734	0.3017	(0.6739, 1.8641)
σ_{Lu}^2	0.6579	0.2578	(0.2926, 1.2740)
ρ_{Cr}	0.7797	0.1580	(0.4070, 0.9898)
ρ_{Le}	0.7713	0.1674	(0.3874, 0.9911)
γ_0	0.9191	0.6165	(-0.1498, 2.2710)
γ_1	-4.9840	0.9955	(-6.9395, -3.0395)

Note: In refers to the intrinsic model, Cv to the convolution model, Cr to the Cressie model, Le to the Leroux model and Lu to the Lu model.

the unstructured spatially effect in the convolution model), all values can be considered close relatives, since their credibility intervals of 95% overlap. The parameters of spatial dependence – main attraction of proper CAR models (Cressie and Leroux) – were estimated to be approximately 0.77, which features a relatively high spatial dependence, although no definitive criteria exist to classify the strength of spatial dependence based on its estimate. For the Lu model, estimates for γ_0 and γ_1 were 0.9191 and -4.9840, respectively. These values lead to a median credibility interval for p_{ij} about (0.36, 0.86). Thus, the information content of the prior distributions of γ_0 and γ_1 , given in Section 2.3, is a price to pay for obtaining additional information on the existing dynamics between regions, to be studied in a possible boundary analysis.

Figure 2 presents the risk of death due to Crohn's Disease and Ulcerative Colitis estimated. According to the maps produced, it is noted that the estimated risks for each microregion are close, for all models, since most of the microregions was classified within the same range.

Table 2 presents the results for the four measurements used to model selection. Note that the CPO does not discriminate any model, assuming the same value for all. On the other hand, the lowest EPD indicates the Leroux model as the best fit and the lowest DIC indicates as the most appropriate the convolution model. The model with the lowest sum of squared residuals is also the convolution, but there is no great difference compared to Cressie and Leroux models. However, none of models has shown definite superiority. This can be analysed visually by Figure 3, which brings the residuals calculated for each model in each microregion of São Paulo. Note that there is a clear similarity between values, indicating that these models actually produced similar estimates.

Either way, taking the convolution model as a reference, higher mortality risks from Crohn's Disease and Ulcerative Colitis were observed in the west (Dracena – 15.34), northwest (Fernandópolis – 4.37, Votuporanga – 3.41, Barretos – 2.87 and Jales – 2.84), northeast (São João da Boa Vista – 2.17) and southeast region (Paraitinga – 2.53). The high risk of Dracena may be considered as an outlier. A specific study of this region is recommended for a possible cause to explain such difference.

Figure 4 summarizes the estimation of the risks, classifying each microregion in three categories. Blank areas are those whose 95% credibility intervals include values smaller than one. Light grey regions have credibility intervals containing the value one. Finally, areas in dark grey

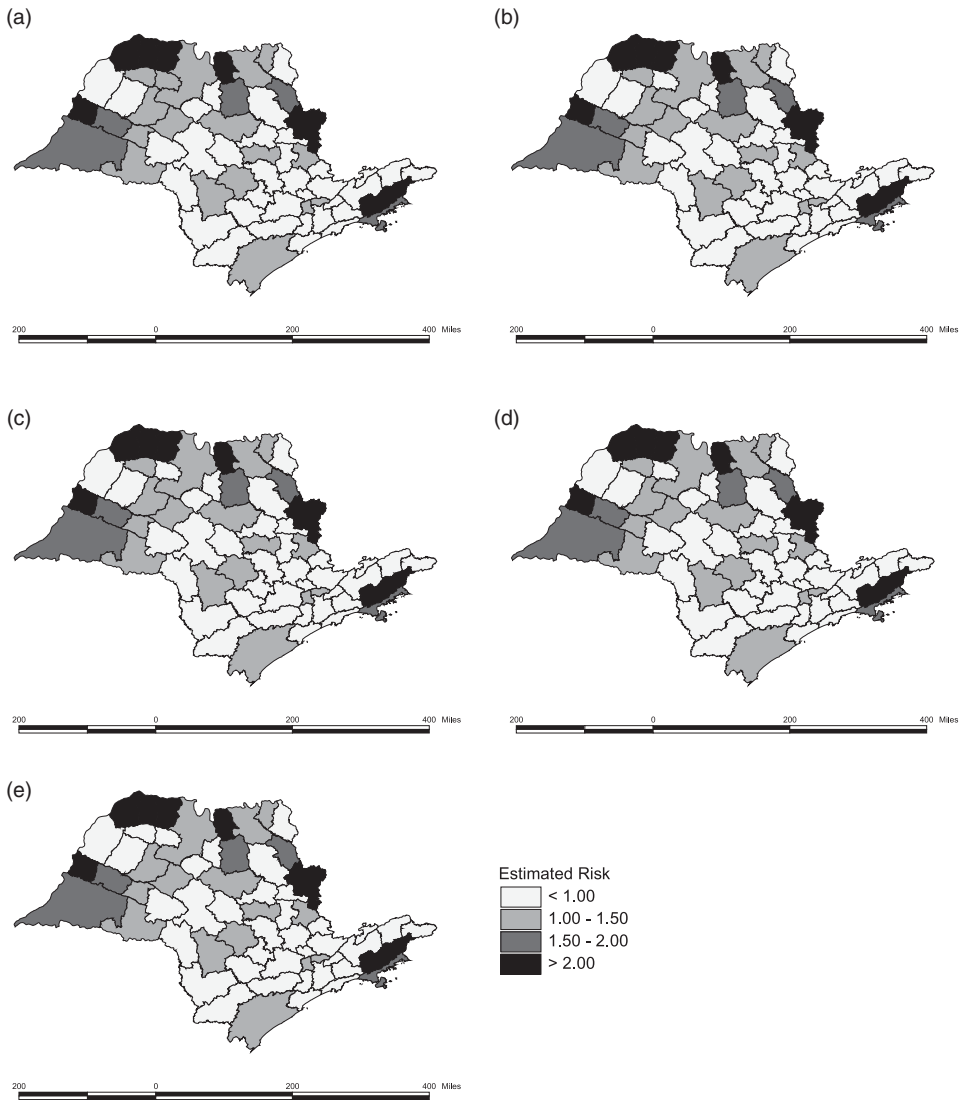


Figure 2. Risk of death due to Crohn’s Disease and Ulcerative Colitis estimated by the intrinsic, convolution, Cressie, Leroux and Lu models for each microregion of the State of São Paulo.

Table 2. CPO, EPD, DIC and RSS provided by the five CAR models for mortality data of Crohn’s disease and Ulcerative Colitis.

Model	Intrinsic	Convolution	Cressie	Leroux	Lu
CPO	4.6	4.6	4.6	4.6	4.6
EPD	10,504.3	10,541.6	10,568.1	10,470.7	10,528.7
DIC	409.1	336.1	464.0	464.6	463.7
RSS	180.3	155.7	159.4	164.6	216.2

Note: Values in bold indicate the best model.

can be considered high-risk regions as long as they have credibility intervals with values greater than one. It may be noted that some groups of high risk are in the western, northern and even in a coastal region São Paulo state.

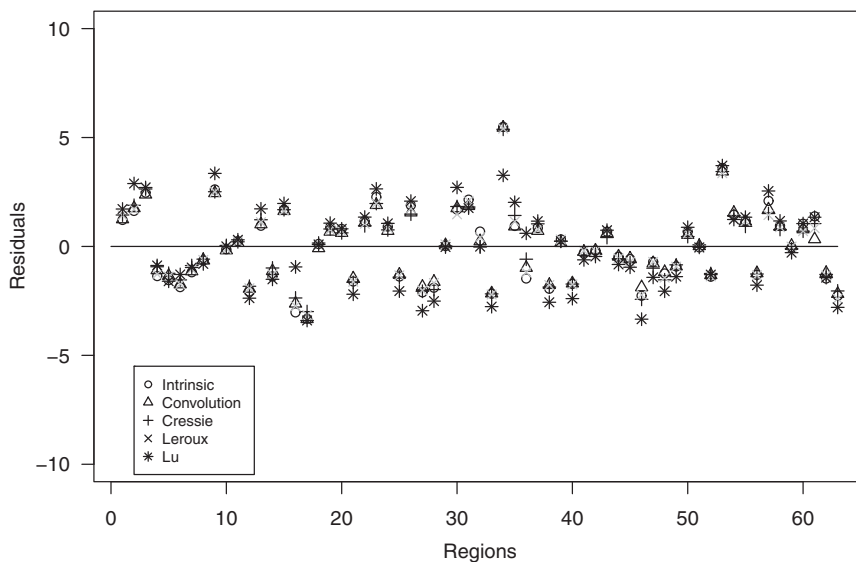


Figure 3. Residuals of five CAR models for each microregion of São Paulo for mortality data of Crohn's Disease and Ulcerative Colitis.

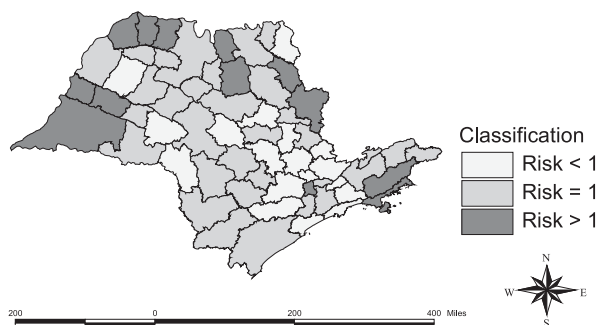


Figure 4. Microregions classified into three groups based on their 95% credibility intervals for data of death by Crohn's and Ulcerative Colitis: low risk (white), within the expected risk (light grey), high risk (dark grey).

Using GeoDa software [29] we also adopted the Moran's I (a descriptive coefficient which measures the spatial autocorrelation) for the random effects estimated by the convolution model and obtained the value of 0.2460. Applying a permutation test, a p -value less than 0.001 was obtained, which features a significant value and corroborates with the capture of spatial structure for the random effects.

4. Simulation study

In order to evaluate the performance of CAR models described in Section 2, a simulation study proposed by Lee [2] was conducted. The data were generated using the map of the 63 microregions of São Paulo as a reference. Intercept term was arbitrarily set at $\alpha = 0.1$. Furthermore, a simulated covariate was included in the data-generating process, with its associated parameter set at $\beta = 1$. Three different scenarios were devised in order to ascertain how the models behave

in the presence of different levels of spatial dependence, including also the case of independence. To generate data with spatial dependence, we used the Matern covariance function.[30] This function can be defined as

$$C(d) = \tau^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu} \frac{d}{\eta} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{d}{\eta} \right),$$

where d is the distance between two regions, η is the amplitude, ν is a smoothing parameter and τ^2 is the marginal variance, with d, η, ν and $\tau > 0$. K_ν is the modified Bessel function of the second kind, which can be written as

$$K_q(z) = \left(\frac{\pi}{2} \right) \frac{I_{-q}(z) - I_q(z)}{\sin(\pi q)}, \quad \text{where } I_q(z) = \left(\frac{z}{2} \right)^q \sum_{p=0}^{\infty} \frac{(z^2/4)^p}{p! \Gamma(q + p + 1)}.$$

From a distance matrix – based on the centroids of each microregion – the Matern function is employed in order to obtain a correlation matrix to be used to generate random effects from a multivariate normal distribution. Using the generated values for the random effects, we can calculate an average rate for each microregion, used as parameters for the Poisson distribution to generate a vector of counts that emulate the number of deaths from the disease.

Finally, the scenarios created for the simulation study were the following:

- Scenario 1: independence. Random effects ϕ were generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and correlation matrix given by Matern class, with parameters $\tau^2 = 1$, $\nu = 2.5$ and negligible amplitude ($\eta = 0.0001$), which, with those specifications, provides an identity matrix.
- Scenario 2: moderate spatial dependence. Random effects ϕ were generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and correlation matrix given by Matern class, with parameters $\tau^2 = 1$, $\nu = 2.5$ and $\eta = 0.15$, which, with these specifications, provide an average correlation between (all) areas of 0.2.
- Scenario 3: strong spatial dependence. Random effects ϕ were generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and correlation matrix given by Matern class, with parameters $\tau^2 = 1$, $\nu = 2.5$ and $\eta = 0.6$, which, with these specifications, provide an average correlation between (all) areas of 0.4.

It is noteworthy that the average correlations defined in the simulation scenarios differ from the dependence parameter ρ of Cressie and Leroux models. The first refers to the average correlations among all regions, including those not adjacent. The second is a dependence parameter restricted to neighbouring areas. Therefore, it is natural to expect that average correlations used to generate the simulation data are markedly smaller than the estimated value for ρ .

For each scenario, we have generated a total of 500 simulated data sets. For each of these data sets, estimation via MCMC was performed with 2000 iterations discarded for burn-in and a total of 8000 subsequent iterations, with a thin equals to 15. Table 3 provides a summary of the results obtained, with statistics as the average value of the parameters α , σ^2 and ρ , plus the mean squared error (MSE) and Bias for the estimated risks. The values of the MSE and the Bias are calculated as

$$\text{MSE}(R) = \sum_{i=1}^n \sum_{j=1}^m \frac{(\hat{R}_{ij} - R_{ij})^2}{m}, \quad \text{and} \quad \text{Bias}(R) = \sum_{i=1}^n \sum_{j=1}^m \frac{(\hat{R}_{ij} - R_{ij})}{m},$$

where $n = 63$ is the number of microregions and $m = 500$ is the total number of iterations performed in the simulation.

Table 3. Results obtained by a simulation study considering three different scenarios.

Metrics	Scenario	Intrinsic	Convolution	Cressie	Leroux	Lu
Mean value of α	1	0.0735	0.0892	0.0921	0.0927	0.1040
	2	0.0903	0.0977	0.1051	0.1018	0.1202
	3	0.0800	0.0825	0.0927	0.0991	0.1132
Mean value of β	1	0.9418	1.0004	0.9853	1.0063	1.0000
	2	0.9498	0.9938	1.0015	1.0028	0.9938
	3	0.9920	1.0009	0.9940	0.9945	0.9884
Mean value of σ^2	1	5.3321	0.6066	5.2243	1.8431	0.9254
	2	4.2561	1.3473	4.5025	2.5127	1.0197
	3	1.0341	0.9929	0.6598	0.5949	1.5975
Mean value of σ_0^2	1	-	0.9190	-	-	-
	2	-	0.6342	-	-	-
	3	-	0.0365	-	-	-
Mean value of ρ	1	-	-	0.2799	0.1745	-
	2	-	-	0.6131	0.4574	-
	3	-	-	0.9679	0.8902	-
MSE(R)	1	4.0605	3.6575	3.6710	3.6583	3.6600
	2	4.0684	3.8178	3.8213	3.8144	3.8176
	3	3.5127	3.4631	3.4389	3.4383	3.5096
Bias(R)	1	-1.1831	-0.0132	0.0208	-0.0073	-0.0120
	2	-1.0858	0.0659	0.0697	0.0711	0.0580
	3	-0.5136	0.0133	-0.0029	-0.0036	-0.0222

All models efficiently estimate the parameters α and β , since, on average, the estimates were set close to its true values ($\alpha = 0.1$ and $\beta = 1$). Regarding the variance parameters, the more accurately models were the convolution, Leroux and Lu. It is noteworthy that the convolution model has two parameters variance, spatially structured and unstructured. It is interesting to note that in the independence scenario, the variance parameter related to the spatially unstructured random effect is higher (0.9190 versus 0.6066). On the other hand, for scenario 3, with the presence of strong spatial dependence, this characteristic is reversed, and now the parameter related to the spatial random effect strongly dominant (0.9929 versus 0.0365). Regarding the ρ parameter, it is noted that it can reflect different levels of spatial correlation. However, it is possible to perceive the existence of a certain overestimation for it, since in the independence scenario, their averages were 0.2799 and 0.1745 for models of Cressie and Leroux, respectively, when in fact these should be located closer to zero.

Figure 5 illustrates the histograms for the estimated values in the simulation for the parameter ρ . In scenario 1, it can be seen a clear positive skewness, indicating that the mean may not be the most appropriate estimator to represent the estimated values. Using the mode, a more robust estimator in the presence of asymmetry, estimates of parameter ρ for Cressie and Leroux models are smaller (0.1941 and 0.0359, respectively), making the results more consistent with the true parameter value. The method used for obtaining the mode was that of Robertson and Cryer.[31]

Considering the MSE(R) and Bias(R) measures, the convolution, Cressie, Leroux and Lu models showed similar results. The only one that diverges with a less satisfactory value is the intrinsic model, especially in scenarios 1 and 2. This is an indication that the absence of the additional parameter ρ slightly penalizes the performance of the intrinsic model for cases of the absence/moderate spatial structure. For all other models, it can be concluded that, despite its peculiarities, advantages and disadvantages, whatever the model used, the risk of mortality will be satisfactorily estimated.

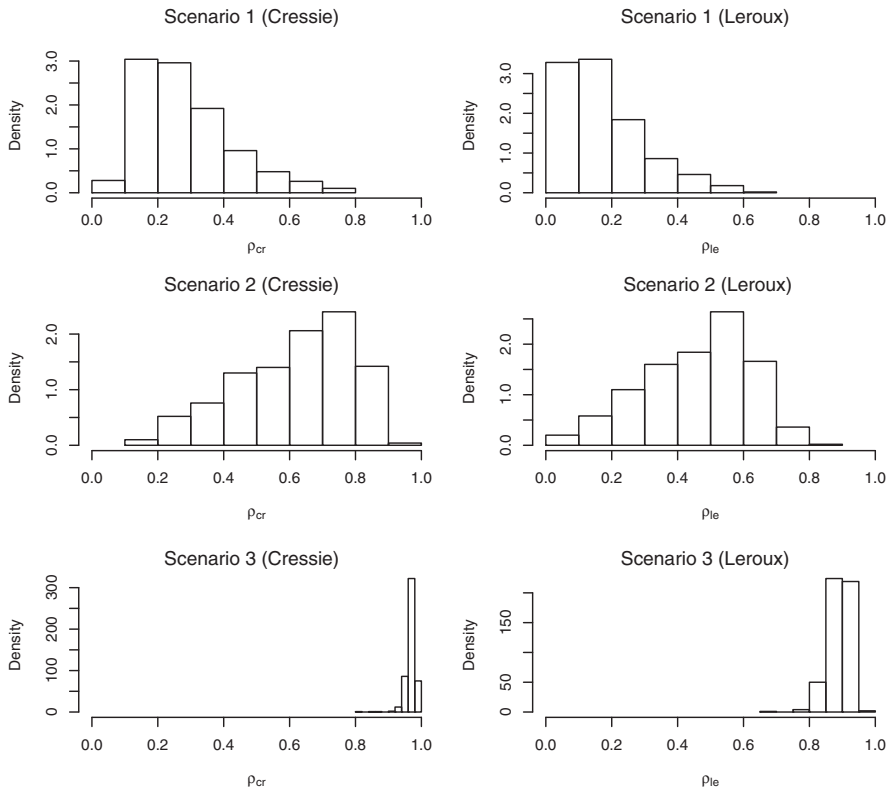


Figure 5. Histograms of the estimated values of ρ_{Cr} and ρ_{Le} at each iteration of the simulation study.

5. Final remarks

In this paper, we explore different conditional autoregressive models with an application in disease mapping. It was revealed the existence of an apparent spatial structure, as evidenced by the estimated values of the spatial dependence parameter in models of Cressie and Leroux, besides the relative variability of the variance parameters of convolution model. The high-risk areas were identified.

All models yielded close estimated risks. This feature has been partially enhanced by a simulation study in which we created three scenarios with different levels of spatial dependence, using a covariance function of the Matern class. This minimizes the need to choose a definitive model that accommodates the random effects, because, concerning the assessment of the risk estimation, all models produced similar results. The exception occurs for the intrinsic model in the scenarios of the absence/moderate spatial structure. The advantages and gains brought by the choice of each model were raised. It is then up to the researcher to define the most relevant feature that meets one’s desires.

Some windows of interest remain open. One is to consider alternative measures to the centroid to represent the information of the region as a covariate in the Lu model. A more plausible alternative is, instead of simply assigning the centre of the area all the information, consider a centroid directed to areas of higher population density. Other covariates could also be adopted in both the logistic model of Lu, as in the general formulation presented in Equation (1). Finally, a boundary analysis can be conducted to the presented data, in order to better understand the nature of neighbourhood relations between the regions of the State of São Paulo – Brazil.

Additionally, further research involving the CAR models can be conducted. Sensitivity studies for γ_0 and γ_1 are of interest to seek alternatives to the informative prior distributions. Finally, performance studies involving generalizations of CAR models described herein for the multivariate generalization could be performed, in order to verify the behaviour of the models for the case of interactions between different types of diseases. The same applies to the temporal generalization.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

R.C. Rampaso  <http://orcid.org/0000-0002-4947-4317>

A.D.P. de Souza  <http://orcid.org/0000-0001-9533-5804>

E.F. Flores  <http://orcid.org/0000-0001-7385-6705>

References

- [1] Stern HS, Cressie N. Posterior predictive model checks for disease mapping models. *Stat Med.* 2000;19:2377–2397.
- [2] Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spatio-Temporal Epidemiol.* 2011;2:79–89.
- [3] Ripley B. *Spatial statistics.* New York: Wiley; 1981.
- [4] Christakos G. On the problem of permissible covariance and variogram models. *Water Resour Res.* 1984;20:251–265.
- [5] Haining R. *Spatial data analysis in the social and environmental sciences.* Cambridge: Cambridge University Press; 1993.
- [6] Wall MM. A close look at the spatial structure implied by the CAR and SAR models. *J Statist Plann Inference.* 2004;121:311–324.
- [7] Alan E Gelfand PJD, Montserrat Fuentes PD. *Handbook of spatial statistics.* New York: Chapman and Hall/CRC Handbooks of Modern Statistical Methods; 2010.
- [8] Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics.* 1987;43:671–681.
- [9] Cressie NAC. *Statistics for spatial data.* Revised ed. New York: Wiley; 1993.
- [10] Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Statist Math.* 1991;43:1–20.
- [11] Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res.* 2005;14:35–59.
- [12] Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika.* 1995;82:733–746.
- [13] Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. *GeoBugs user manual.* Version 1.2. September 2004. Available from: <http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/geobugs12manual.pdf>.
- [14] Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran ME, Berry D, editors. *Statistical models in epidemiology, the environment and clinical trials.* New York: Springer; 1999. p. 179–191.
- [15] Breslow NE. Extra-Poisson variation in log-linear models. *Appl Statist.* 1984;33:38–44.
- [16] Rodrigues EC, Assunção R. Bayesian spatial models with a mixture neighborhood structure. *J Multivariate Anal.* 2012;109:88–102.
- [17] Eberly LE, Carlin BP, and others. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Stat Med.* 2000;19:2279–2294.
- [18] Rodrigues EC. *Inferindo a Estrutura de Vizinhança em Modelos Bayesianos Espaciais.* Master's thesis, Pós-Graduação em Estatística, UFMG, 2011.
- [19] Banerjee S, Gelfand AE, Carlin BP. *Hierarchical modeling and analysis for spatial data.* London: CRC Press; 2003.
- [20] Lu H, Reilly CS, Banerjee S, Carlin BP. Bayesian areal wobble via adjacency modeling. *Environ Ecol Stat.* 2007;14:433–452.
- [21] Womble WH. Differential systematics. *Science.* 1951;114:315–322.
- [22] Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* 2006;1:515–534.
- [23] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. *Stat Med.* 2009;28:3049–3067.
- [24] Ghosh P, Huang L, Yu B, Tiwari RC. Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data. *Comput Statist Data Anal.* 2009;53:4073–4082.

- [25] Chen MH, Shao QM, Ibrahim JG. Monte Carlo methods in Bayesian computation. Springer Series in Statistics, New York: Springer-Verlag; 2000.
- [26] Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss approach. *Biometrika*. 1998;85:1–11.
- [27] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B (Stat Methodol)*. 2002;64:583–639.
- [28] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- [29] Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geographical Anal*. 2006;38:5–22.
- [30] Minasny B, McBratney AB. The Matérn function as a general model for soil variograms. *Geoderma*. 2005;128:192–207.
- [31] Robertson T, Cryer JD. An iterative procedure for estimating the mode. *J Amer Statist Assoc*. 1974;69:1012–1016.