

# On the use of adaptive spatial weight matrices from disease mapping multivariate analyses\*

Corpas-Burgos, Francisca  
*Health Inequalities Research Area*  
*FISABIO Foundation, Valencia, Spain.*

Martinez-Beneito, Miguel A.  
*Department of Statistics and Operations Research,*  
*University of Valencia, Valencia, Spain.*

## Abstract

Conditional autoregressive distributions are commonly used to model spatial dependence between nearby geographic units in disease mapping studies. These distributions induce spatial dependence by means of a spatial weights matrix that quantifies the strength of dependence between any two neighboring spatial units. The most common procedure for defining that spatial weights matrix is using an adjacency criterion. In that case, all pairs of spatial units with adjacent borders are given the same weight (typically 1) and the remaining non-adjacent units are assigned a weight of 0. However, assuming all spatial neighbors in a model to be equally influential could be possibly a too rigid or inappropriate assumption. In this paper, we propose several adaptive conditional autoregressive distributions in which the spatial weights for adjacent areas are random variables, and we discuss their use in spatial disease mapping models. We will introduce our proposal in a multivariate context so that the spatial dependence structure between spatial units is shared and estimated from a sufficiently large set of mortality causes. As we will see, this is a key aspect for making inference on the spatial dependence structure. We show that our adaptive modeling proposal provides more flexible and accurate mortality risk estimates than traditional proposals in which spatial weights for neighboring areas are fixed to a common value.

## 1 Introduction

Disease mapping has attracted considerable attention over the last three decades (Lawson, 2018; Martinez-Beneito and Botella Rocamora, 2019). This area of

---

\*This is a preprint version of the paper: Corpas-Burgos, F; Martínez-Beneito, MA. On the use of adaptive spatial weight matrices from disease mapping multivariate analyses. *Stochastic Environmental Research and Risk Assessment*, 2020, 34, 531-544.

research pursues the study of the geographical distribution of health-related events, such as mortality from, or incidence of diseases, aggregated over areal units, in order to identify mainly those locations which show higher risks. In disease mapping problems, the units of study usually considered are as small as possible, which can lead to what are known as small areas estimation problems. As a consequence, many modeling proposals have been formulated in order to deal with this problem and thereby derive reliable risks estimates. Most of these models consider dependence among nearby spatial units, assuming them to show similar risks. Therefore the spatial dependence hypothesis is the main key to improving risks estimates in disease mapping studies.

A large number of disease mapping models have been proposed, most of them following a Bayesian approach; see Besag et al. (1991) or Leroux et al. (1999) for two of the most frequently used models in applied studies. These proposals are frequently specified as generalized linear models that incorporate spatial dependence between nearby geographical units through random effects following some spatial prior distribution. Although some other spatial modeling tools have been also used (Adín et al., 2017), the most popular spatial prior distributions in disease mapping models belong to the family of Conditional Autoregressive (CAR) distributions (Besag, 1974; Besag et al., 1991), also known as Gaussian Markov Random Fields (GMRF) (Rue and Held, 2005). CAR distributions induce spatial dependence by considering a schematic neighborhood structure which accounts for the geographical arrangement of the spatial units. That neighborhood structure is summarized by means of a spatial weights matrix quantifying the relative influence that the random effects of the geographical units have on each other, so those weights should reflect the strength of the dependence between any pair of spatial units. Moreover, that weights matrix is usually sparse, reflecting an implicit Markovian assumption which considers the conditional distribution of any random effect, given its neighbors, independent of the random effects in any other spatial location.

Different proposals of spatial weights matrices for CAR distributions have been made in the literature. By far, the most common procedure is using an adjacency criterion for defining that matrix. In that case all pairs of spatial units with adjacent geographical borders are given the same weight, typically 1, and the remaining non-adjacent units are assigned a weight of 0, reflecting independence given the remaining spatial units (Besag et al., 1991). As pointed out by Raftery and Banfield (1991), this choice could be sensible for regular lattices but less so for irregular lattices such as those typically used in disease mapping problems. Weighted versions of this common choice also exist which are available in some common Bayesian inference packages, such as `WinBUGS` or `OpenBUGS`; while some others, such as `INLA`, do not allow for this option. This allows the strength of the dependence of nearby pairs of units to be modulated, thereby allowing each neighboring pair of units to show a different strength. However, the CAR distributions in `WinBUGS` or `OpenBUGS` do not allow those weights to be estimated as variables within a model; on the contrary, they have to be supplied as constants to the corresponding model. Best et al. (1999) and Earnest et al. (2007), for example, propose the use of weights matrices with dif-

ferent weights which are a function of the geographic distance between spatial units (usually, the Euclidean distance between their centroids); in this manner random effects of closer geographic units will show stronger dependence. However, it could happen that geography is not necessarily the main determinant of dependence between units; thus areas with similar values of certain covariates, for example, would take similar risks estimates in general even though they are distant. In this regard, Kuhnert (2003) defines the weights matrix of the random effects as a function of the absolute difference between the values of some covariate for the spatial units. Likewise, Earnest et al. (2007) define the weights as a function of both the geographical distance and their similitude in terms of some covariate. A comparison of models of this kind is undertaken in Duncan et al. (2017).

Despite their interest to researchers, the use of the weight matrices above shows some limitations. Firstly, unweighted adjacency-based matrices do not have clear support beyond their simplicity and convenience. In the end, assuming all spatial neighbors in a model to be equally influential is an arbitrary assumption that should be checked in some way. Nevertheless, the mentioned convenience of that choice has led most disease mapping practitioners to accept and use that matrix, without further justification, and to avoid questioning that assumption. On the other hand, the use of functions of geographic distance for setting weights matrices assumes equal weights for all locations which are equally distant, which could be somewhat simplistic for some settings (consider regions with mountains, rivers or other barriers). Additionally, those distance weighted proposals assume parametric relationships between distances and weights, which could also be rigid or sometimes inappropriate. Finally, the definition of weights as a function of some covariate poses an additional problem since the corresponding covariate may not always be available for all locations. Therefore, the requirements for this option are higher than for pure geometric criteria.

The objective of this work is to propose a procedure for estimating the spatial weights matrix in disease mapping studies solving the issues above. Specifically, we focus on the barely explored adaptive CAR distributions which consider the weights of the spatial weights matrix as additional random variables in the model. Some works can be found in the literature, such as MacNab et al. (2006); Brezger et al. (2007); Lu et al. (2007); Congdon (2008); Ma et al. (2010), that follow this approach. Our proposal, in contrast to the previous works, estimates a common weights matrix from the joint study of several diseases, which would, presumably, capture the different dependence strengths shown by the neighboring spatial units in the region of study. As we will see, that multivariate feature of our proposal will be a key aspect for its success. The multivariate estimated weights matrix could be subsequently used in future studies on that same region of study. In principle the enhanced weights structure estimated for that region would allow improved risk estimates to be derived incorporating the dependence structure shown by some set of diseases in that region. That spatial structure should reflect physical/social barriers, data artifacts, geographical/geometrical/social features etc., which would be recommendable to consider

in subsequent spatial analyses on that same region.

This paper is structured as follows. Section 2 introduces some traditional spatial modeling proposals widely used in disease mapping studies and makes a brief review of the main adaptive CAR models already proposed in the literature. Section 3 describes our multivariate modeling proposal for the weights matrix of CAR distributions. Section 4 illustrates how the developments proposed at Section 3 can be used for estimating the spatial dependence structure in a real setting and how that estimation improves subsequent analyses in comparison to studies with unweighted dependence matrices. Finally, Section 5, discusses some results and conclusions drawn from this study.

## 2 Some modeling proposals in disease mapping

### 2.1 Some popular disease mapping models

Disease mapping studies consider regions of interest discretely divided into  $I$  spatial units, generally of small size, such as census tracts or municipalities. The main aim of these studies is to determine the geographical distribution of the risks for some disease for these spatial units. The collection of observed cases per spatial unit are jointly denoted by  $\mathbf{O} = (O_1, \dots, O_I)'$ , where  $O_i$  denotes the number of observed cases in the  $i$ -th unit. Typically, disease mapping models assume:

$$O_i \sim \text{Poisson}(E_i R_i), \quad i = 1, \dots, I,$$

where  $\mathbf{E} = (E_1, \dots, E_I)'$  contains the number of expected cases per spatial unit for the corresponding disease and  $\mathbf{R} = (R_1, \dots, R_I)'$  is the collection of location specific risks that we would want to estimate. Typically, the log-risks are modeled as:

$$\log(R_i) = \mu + \mathbf{z}_i^T \boldsymbol{\beta} + \eta_i, \quad (1)$$

where  $\mu$  is an intercept,  $\mathbf{z}_i$  is a vector of covariates, with  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  being its vector of associated parameters, and  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_I)'$  a vector of random effects. The random effects  $\boldsymbol{\eta}$  are introduced to explain the variability that cannot be explained by the covariates and  $\boldsymbol{\eta}$  is typically assumed to show spatial dependence since that residual variability could easily exhibit that feature. From now on, for simplicity, we will assume that no covariates are available and the log-risks are simply modeled as the sum of the intercept and the set of random effects.

The random effects vector  $\boldsymbol{\eta}$  is habitually modeled by using spatially correlated CAR prior distributions. A particularly popular case of CAR prior distribution is the Intrinsic CAR (Besag et al., 1991) distribution (ICAR from now on), which for a vector  $\boldsymbol{\phi}$  may be defined by the following set of  $I$  univariate conditional distributions:

$$\phi_i | \boldsymbol{\phi}_{-i}, \sigma_\phi^2 \sim N \left( \frac{1}{w_{i+}} \sum_{k=1}^I w_{ik} \phi_k, \frac{\sigma_\phi^2}{w_{i+}} \right), \quad i = 1, \dots, I. \quad (2)$$

In this expression, the subindex in  $\phi_{-i}$  denotes all the terms in  $\phi$  excepting its  $i$ -th component,  $w_{ik}$  weighs the contribution of the  $k$ -th random effect to the mean of  $\phi_i$ ,  $w_{i+} = \sum_{k=1}^I w_{ik}$  and  $\sigma_\phi^2$  is a variance parameter. These conditional distributions can be shown (Besag, 1974) to yield the following joint distribution for  $\phi$ :

$$\phi | \sigma_\phi^2 \sim N_I(\mathbf{0}, \sigma_\phi^2(\mathbf{D} - \mathbf{W})^-),$$

where  $\mathbf{D} = \text{diag}(w_{1+}, \dots, w_{I+})$ ,  $\mathbf{W} = (w_{ik})_{i,k=1}^I$  and the superindex in  $(\mathbf{D} - \mathbf{W})^-$  denotes the Moore-Penrose inverse of  $\mathbf{D} - \mathbf{W}$ . Covariance between elements of  $\phi$  is determined by the spatial weights matrix  $\mathbf{W}$ , whose elements  $w_{ik}$  are typically non-zero if areas  $(i, k)$  are considered neighbors and zero otherwise. Therefore, if two areas are considered neighbors, their random effects are conditionally dependent, while random effects of non-neighboring areas are conditionally independent. As mentioned previously, the most common approach is to assume that areas  $(i, k)$  are neighbors if they share a common border (adjacency) and in that case set  $w_{ik} = 1$  for all neighboring pairs of units  $(i, k)$ . In that case, the conditional distributions above reduce to simply:

$$\phi_i | \phi_{-i}, \sigma_\phi^2 \sim N\left(\frac{1}{n_i} \sum_{k \sim i} \phi_k, \frac{\sigma_\phi^2}{n_i}\right), \quad i = 1, \dots, I, \quad (3)$$

where  $n_i$  stands for the number of neighboring areas of unit  $i$  and the subindex  $k \sim i$  denotes all those units  $k$  which are neighbors of  $i$ . Now, the conditional mean of  $\phi_i$  is equal to the raw (unweighted) mean of the random effects in its neighboring areas and its conditional variance is inversely proportional to the number of neighbors  $n_i$ .

One of the most popular specifications for  $\boldsymbol{\eta}$  in disease mapping studies is that introduced in Besag et al. (1991) (BYM from now on). In this proposal, the random effects vector  $\boldsymbol{\eta}$  is considered to be the sum of two vectors of random effects  $\boldsymbol{\eta} = \boldsymbol{\phi} + \boldsymbol{\theta}$ . The term  $\boldsymbol{\phi}$ , which follows an ICAR distribution as just introduced, will be responsible for inducing spatial dependence on  $\mathbf{R}$  and accounts for those risk factors of regional scope which take effect on several contiguous spatial units, making them in principle similar. The second term,  $\boldsymbol{\theta}$ , whose components follow independent Normal distributions of mean zero and common variance  $\sigma_\theta^2$ , accounts for risk factors of very limited geographical scope that take an effect just on isolated areal units, making their risks different to those of their surrounding units. Thus, this second term induces additional unstructured variability in  $\boldsymbol{\eta}$ . The amount of spatial/unstructured variability in  $\mathbf{R}$  depends on the balance between  $\sigma_\phi$  and  $\sigma_\theta$ , which is determined by the model/data itself. If the former has higher (respectively lower) values, in comparison to the latter, the final pattern will show substantial spatial dependence (respectively independence).

A second popular CAR prior distribution for inducing spatial correlation on the random effects vector  $\boldsymbol{\eta}$  in Expression (1) is that introduced in Leroux et al. (1999). In contrast to the BYM model,  $\boldsymbol{\eta}$  in this alternative proposal is not the sum of two additional components. In this case, the determination of the

amount of spatial/unstructured variability is controlled by a spatial correlation parameter  $\rho \in [0, 1]$  so that the special case of  $\rho = 0$  simplifies to a model with independent random effects and  $\rho = 1$  corresponds to the ICAR distribution above. All intermediate values of  $\rho \in (0, 1)$  induce patterns mixing both sources of dependence. Specifically, for the Leroux et al. proposal, the prior conditional distributions corresponding to  $\eta_i$  are given by:

$$\eta_i | \boldsymbol{\eta}_{-i}, \rho, \sigma_\eta^2 \sim N \left( \frac{\rho}{\rho w_{i+} + 1 - \rho} \sum_{k=1}^I w_{ik} \eta_k, \frac{\sigma_\eta^2}{\rho w_{i+} + 1 - \rho} \right), \quad i = 1, \dots, I.$$

Note the obvious coincidence of this proposal with a weighted CAR distribution for  $\rho = 1$  and with a heterogeneous Normal distribution for  $\rho = 0$ . For the usual assumption of  $w_{ik} = 1$  for adjacent spatial units, and 0 otherwise, the Leroux et al. proposal reduces to:

$$\eta_i | \boldsymbol{\eta}_{-i}, \rho, \sigma_\eta^2 \sim N \left( \frac{\rho}{\rho n_i + 1 - \rho} \sum_{k \sim i} \eta_k, \frac{\sigma_\eta^2}{\rho n_i + 1 - \rho} \right), \quad i = 1, \dots, I.$$

In the same manner as for the ICAR distribution, these conditional distributions yield a joint Normal distribution, specifically:

$$\boldsymbol{\eta} | \rho, \sigma_\eta^2 \sim N_I(\mathbf{0}, \sigma_\eta^2 ((1 - \rho) \mathbf{I}_I + \rho(\mathbf{D} - \mathbf{W}))^{-1}), \quad i = 1, \dots, I,$$

where, as for BYM,  $\mathbf{W} = (w_{ik})_{i,k=1}^I$  denotes the spatial weights matrix considered.

As described in the introduction, setting the same weights  $w_{ik} = 1$  to all the random effects of adjacent locations in CAR distributions could be an inappropriate or rigid assumption. This makes all neighboring regions equally influential on any particular risk, which may not correspond to reality. In order to solve this limitation, models with alternative stochastic weight matrices have been proposed and are reviewed in the next subsection.

## 2.2 Adaptive CAR distributions

A few CAR models with adaptive weights matrices can be found in the Bayesian disease mapping literature. The goal of these proposals is to model spatial correlation through the fitting of an stochastic spatial weights matrix  $\mathbf{W}$ . This approach is undertaken within Bayesian hierarchical models where the corresponding CAR distributed random effects are defined as  $\boldsymbol{\eta} | \mathbf{W}, \sigma_\eta^2, \dots$  and an additional layer is considered in the model for estimating the elements in  $\mathbf{W}$ . Next, we briefly summarize several of the contributions in this area. There is obviously a huge body of literature proposing the stochastic modeling of variances or covariances between variables in many different contexts: state space models (Carter and Kohn, 1996), function estimation (Lang et al., 2002), etc. Nevertheless, we will restrict the review below just to CAR spatial models in order to focus on the particular topic that we are discussing where the variance matrix is defined as a function of a particular weights matrix  $\mathbf{W}$ .

*Wombling* (Lu and Carlin, 2005; Lu et al., 2007; Ma and Carlin, 2007; Ma et al., 2010) would be a first attempt of stochastic modeling of the spatial weights matrix  $\mathbf{W}$  in CAR distributions. Specifically, *Wombling* assumes the cells of  $\mathbf{W}$  as binary stochastic values, which are modeled as Bernoulli distributions. The probability of  $w_{ij} = 1$  for any pair of spatial units  $(i, j)$  could be modeled by means of a logistic regression as a function of some covariates (Lu et al., 2007), such as the adjacency matrix of the area of study or some other related quantity. Obviously, the number of elements in that logistic regression will increase quadratically as a function of the number of spatial units in the study, which could be a problem for large lattices. In addition, a large collection of sensible covariates would be required under this approach in order to define a rich enough spatial weights matrix. More flexible alternatives are also considered for estimating  $\mathbf{W}$  within the *Wombling* approach, although in this case only the weights of the cells corresponding to adjacent elements in the lattice are estimated. In this case, since  $w_{ij}$  are modeled as Bernoulli variables for adjacent units, this procedure will prune the adjacency graph originally considered. Ma et al. (2010) proposes a spatial Ising model (see for example Geman and Geman (1984) or Green and Richardson (2002)), which favors contiguous edges in the graph (those sharing one of their nodes) to take the same values in the weights matrix. This proposal has also been applied to multivariate data sets, for the joint study of three diseases, as in Ma and Carlin (2007). Although this proposal seems much more flexible than the naive use of covariates for modeling  $P(w_{ij} = 1)$ , the binary treatment of the elements of the weights matrix in *Wombling* studies seems somewhat restrictive. Moreover, the use of Ising models for estimating the non-zero cells of  $\mathbf{W}$  induces important computational problems (Ma and Carlin, 2007) for estimating the penalizing parameter of that model, at least in the multivariate case. This forces this parameter to be fixed/tuned according to previous runs of the models. However, as reported by the authors, the fit of this model even for a fixed penalizing parameter becomes challenging for medium/large lattices.

On the other hand, MacNab et al. (2006) and Congdon (2008) consider adaptive versions of the Leroux et al. CAR prior distribution. This approach could be also used for estimating weights matrices in CAR distributions as will become evident in the next section. These proposals allow the spatial correlation parameter  $\rho$  of Leroux et al. to vary for each geographical unit. Specifically, MacNab et al. (2006) propose defining an unweighted (adjacency-based) spatial process as the following set of conditional distributions:

$$\eta_i | \boldsymbol{\eta}_{-i}, \boldsymbol{\rho}, \sigma_\eta^2 \sim N \left( \frac{\rho_i}{\rho_i n_i + 1 - \rho_i} \sum_{k \sim i} \eta_k, \frac{\sigma_\eta^2}{\rho_i n_i + 1 - \rho_i} \right), \quad i = 1, \dots, I.$$

The problem with this proposal is that this set of conditional distributions does not yield a valid CAR prior distribution since the symmetry condition (Besag and Kooperberg, 1995), necessary for  $\boldsymbol{\eta}$  to have a symmetric covariance matrix, does not hold in this case. Interestingly, regarding the spatial correlation parameters  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_I)$ , MacNab et al. (2006) mention that 'the analysis

showed very little prior-to-posterior updating for the  $\rho_j$ s, an indication that the data did not provide enough information for useful posterior inference'. Congdon (2008), also in the unweighted case for simplicity, proposes the following set of conditional distributions:

$$\eta_i | \boldsymbol{\eta}_{-i}, \boldsymbol{\rho}, \sigma_\eta^2 \sim N \left( \frac{\rho_i}{\rho_i n_i + 1 - \rho_i} \sum_{k \sim i} \rho_k \eta_k, \frac{\sigma_\eta^2}{\rho_i n_i + 1 - \rho_i} \right), \quad i = 1, \dots, I,$$

which fulfills the mentioned symmetry condition. Although this proposal is supposed to extend the Leroux et al. CAR distribution to having different correlation parameters  $\rho_i, i = 1, \dots, I$ , strikingly it does not coincide with that proposal when all those  $\rho_i$  take a single common value  $\rho$ . Moreover, this process yields the following joint covariance matrix:

$$\sigma_\eta^2 (\text{diag}(\mathbf{1}_I - \boldsymbol{\rho}) \mathbf{I}_I + \text{diag}(\boldsymbol{\rho})(\mathbf{D} - \mathbf{W} \text{diag}(\boldsymbol{\rho})))^{-1}.$$

In this case the covariance matrix is not a combination of  $\mathbf{I}_I$  and  $\mathbf{D} - \mathbf{W}$  since this latest term is replaced by  $\mathbf{D} - \mathbf{W} \text{diag}(\boldsymbol{\rho})$ . As a consequence, this proposal does not seem such a straightforward generalization of the Leroux covariance matrix. Congdon (2008) suggests several different prior distributions for the components of  $\boldsymbol{\rho}$ , such as beta, probit-normal or logit-normal distributions which allow further modeling of these variables. This adaptive CAR distribution is proposed and used in the univariate context where a single spatial pattern is studied.

Brezger et al. (2007) propose an alternative adaptive model which makes it possible to make inference on the spatial weights matrix of ICAR prior distributions. The context of this paper is spatio-temporal modeling in human brain mapping, but their ideas could be also useful for regular disease mapping studies. Brezger et al. uses an ICAR as prior distribution for the coefficients of some basis of functions modeling the time trend for a set of brain voxels; a different ICAR distribution is used for each element in that basis. Under this approach, the cells of the weights matrices for those ICAR random effects are considered equal to 0 for all non-adjacent pairs of units and are modeled as positive random variables, following an informative Gamma(0.5,0.5) prior distribution for the adjacent areas. Posterior sensitivity to that informative prior distribution is not assessed in the paper. Additionally, the likelihood for the Brezger et al. proposal is Normal so the applicability of their model and results to regular disease mapping studies, usually with Poisson or binomial likelihoods, should be further explored.

### 3 A new adaptive CAR distribution and its use in multivariate models

As an alternative to the commonly used adjacency criterion, which considers all the weights in  $\mathbf{W}$  as fixed binary quantities, we propose to model the spatial



weights as random variables within the model so as to allow variability between them. We begin by describing two proposals in the simplest univariate case for both ICAR and Leroux et al. spatial distributions. Subsequently, we will describe their equivalents in the context of the multivariate study of several diseases since, as we will argue, this is the appropriate context where adaptive proposals should be implemented.

### 3.1 Univariate case

We start first by introducing the estimation of spatial weights matrices for ICAR distributions. Let  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_I)'$  be a random effects vector with ICAR distribution, that is,  $\boldsymbol{\phi} | \sigma_\phi^2 \sim N_I(\mathbf{0}, \sigma_\phi^2(\mathbf{D} - \mathbf{W})^-)$ . For this expression we will assume that  $\mathbf{D}$  and  $\mathbf{W}$  are defined according to adjacency between spatial units, i. e.  $\mathbf{D} = \text{diag}(n_1, \dots, n_I)$  for  $n_i$  the number of neighbors of unit  $i$  and  $\mathbf{W} = (w_{ik})$  where  $w_{ik} = 1$  if  $(i, k)$  are adjacent units and 0 otherwise.

Let us now consider a random vector  $\mathbf{c} = (c_1, \dots, c_I)'$  of positive values, a new spatial weights matrix  $\mathbf{W}^*(\mathbf{c}) = \text{diag}(\mathbf{c})^{1/2} \mathbf{W} \text{diag}(\mathbf{c})^{1/2}$  and  $\mathbf{D}^* = \text{diag}(w_{1+}^*, \dots, w_{I+}^*)$ . With this, we propose the following adaptive CAR prior distribution:

$$\begin{aligned} \boldsymbol{\phi} | \mathbf{c}, \sigma_\phi^2 &\sim N_I(\mathbf{0}, \sigma_\phi^2(\mathbf{D}^* - \mathbf{W}^*(\mathbf{c}))^-) \\ c_i &\sim \text{Gamma}(\alpha, \alpha). \end{aligned}$$

The elements of the vector  $\mathbf{c}$  are assumed to be positive since the non-zero weights of the new spatial weights matrix  $\mathbf{W}^*$  are  $w_{ij}^* = (c_i c_j)^{1/2}$  so, in this manner they will all be well defined and positive. Accordingly, we have used a Gamma prior distribution for its elements, which seems a natural choice. The Gamma distribution considered has mean 1, in accordance with the value of the non-zero cells of  $\mathbf{W}$  when an adjacency criterion is considered. Thus,  $\mathbf{W}^*(\mathbf{c})$  will be on average equal to  $\mathbf{W}$ , although its non-zero weights will not necessarily have to be equal to 1. Hence the new adaptive distribution will be more flexible than the regular ICAR distribution. Note that, as defined, the (a priori) standard deviation of any element of  $\mathbf{c}$  is equal to  $\alpha^{-0.5}$ , which could guide us to set a prior distribution for this parameter. In fact, we have considered a uniform prior distribution on  $\alpha^{-0.5}$ , with lower and upper limits intended to make it vague, in order to complete the hierarchical structure above.

Alternatively, the definition of the adaptive ICAR distribution above could be restated as a set of conditional distributions  $\phi_i | \boldsymbol{\phi}_{-i}, \mathbf{c}, \sigma_\phi^2$ ,  $i = 1, \dots, I$ , of mean

$$E[\phi_i | \boldsymbol{\phi}_{-i}, \mathbf{c}, \sigma_\phi^2] = \frac{1}{w_{i+}^*} \sum_{k=1}^I w_{ik}^* \phi_k = \frac{c_i^{1/2} \sum_{k \sim i} c_k^{1/2} \phi_k}{c_i^{1/2} \sum_{k \sim i} c_k^{1/2}} = \frac{\sum_{k \sim i} c_k^{1/2} \phi_k}{\sum_{k \sim i} c_k^{1/2}} \quad (4)$$

and variance

$$\text{Var}[\phi_i | \boldsymbol{\phi}_{-i}, \mathbf{c}, \sigma_\phi^2] = \frac{\sigma_\phi^2}{w_{i+}^*} = \frac{\sigma_\phi^2}{c_i^{1/2} \sum_{k \sim i} c_k^{1/2}}. \quad (5)$$

These two expressions provide some quite valuable insights about on the model proposed. The expected value in Expression (4) is just a weighted mean of the random effects for the corresponding neighbors. The weights in that expression are given by the vector  $\mathbf{c}$ , thus if  $c_i$  had a low value for some  $i$ , that region will have a low contribution to the means of its surroundings units. Additionally, Expression (5) suggests that if  $c_i$  is low, then the conditional variance of  $\phi_i$  will be in contrast high. Thus, if  $c_i$  was low, these two expressions suggest that it is as if spatial unit  $i$  was 'disconnected' from its spatial neighbors, since  $\phi_i$  will be less influential on them and will have higher variance, allowing it to move independently from the rest of the units. Conversely, if  $c_i$  was high, unit  $i$  will become more influential on its neighbors and will take a value in close agreement with them. Therefore, in some manner, the adaptive ICAR distribution would impose a tighter dependence between this unit and its neighbors.

Besides the enhanced interpretation just made, the conditional statement of the adaptive ICAR distribution above allows its implementation in conventional Bayesian software packages such as WinBUGS, JAGS ... Additional care should be taken for the adaptive ICAR distribution since sum-to-zero restrictions are, in general, imposed on ICAR distributions in order to solve the rank-deficiency of its precision matrix (Besag and Kooperberg, 1995). This can be done in practice, in a computationally convenient manner, by imposing  $\sum_i \phi_i \sim N(0, \epsilon)$  for some small value  $\epsilon$ . Details of the coding of this constraint for the adaptive ICAR distribution can be found in the supplementary material of the paper (RCode.pdf file).

In the case of the Leroux et al. model  $\phi$  is distributed as  $\phi|\rho, \sigma_\phi^2 \sim N_I(\mathbf{0}, \sigma_\phi^2((1-\rho)\mathbf{I}_I + \rho(\mathbf{D} - \mathbf{W}))^{-1})$ . Following the development above, several adaptive versions of the Leroux et al. distribution could be made. For example, let us assume  $\phi|\rho, \mathbf{c}, \sigma_\phi^2 \sim N_I(\mathbf{0}, \sigma_\phi^2((1-\rho)\text{diag}(\mathbf{c}^{1/2}) + \rho(\mathbf{D}^* - \mathbf{W}^*(\mathbf{c})))^{-1})$ , where  $\mathbf{D}^*$  and  $\mathbf{W}^*(\mathbf{c})$  are defined as for the adaptive ICAR distribution. In this manner, for  $\rho = 1$  this distribution would be equivalent to an adaptive ICAR distribution, while for  $\rho = 0$  it would yield a collection of independent Normal random effects with adaptive (heteroscedastic) variance. If preferred, an alternative formulation of adaptive Leroux distribution could be derived as  $\phi|\rho, \mathbf{c}, \sigma_\phi^2 \sim N_I(\mathbf{0}, \sigma_\phi^2((1-\rho)\mathbf{I}_I + \rho(\mathbf{D}^* - \mathbf{W}^*(\mathbf{c})))^{-1})$  which for  $\rho = 0$  yields non-adaptive (homoscedastic) independent random effects. Nevertheless, we will focus in the first of these options as it seems more flexible and appealing, from our perspective. For that proposal, the conditional mean and variance of the random effect  $\phi_i$  can be expressed as:

$$\begin{aligned} E[\phi_i|\phi_{-i}, \rho, \mathbf{c}, \sigma_\phi^2] &= \frac{\rho}{(1-\rho)c_i^{1/2} + \rho w_{i+}^*} \sum_{k=1}^I w_{ik}^* \phi_k \\ &= \frac{\rho c_i^{1/2}}{(1-\rho)c_i^{1/2} + \rho c_i^{1/2} \sum_{k \sim i} c_k^{1/2}} \sum_{k \sim i} c_k^{1/2} \phi_k \\ &= \frac{\rho}{1-\rho + \rho \sum_{k \sim i} c_k^{1/2}} \sum_{k \sim i} c_k^{1/2} \phi_k \end{aligned}$$

and

$$\text{Var}[\phi_i | \phi_{-i}, \rho, \mathbf{c}, \sigma_\phi^2] = \frac{\sigma_\phi^2}{(1 - \rho)c_i^{1/2} + \rho w_{i+}^*} = \frac{\sigma_\phi^2}{c_i^{1/2}(1 - \rho + \rho \sum_{k \sim i} c_k^{1/2})}.$$

Once again, these conditional expressions allow the adaptive Leroux CAR distribution to be implemented in conventional Bayesian software, such as WinBUGS. Note that similar, although somewhat different, adaptive CAR distributions have been already proposed in MacNab (2018). Nevertheless, those proposals had as a goal the formulation of more general (adaptive) CAR distributions. Our goal will now be to estimate spatial weight matrices  $\mathbf{W}^*(\mathbf{c})$  suitable to be used in subsequent spatial analyses in that same area of study. It is hoped that  $\mathbf{W}^*(\mathbf{c})$  will capture the geometric/demographic/geographic features of the region of study which make some neighboring units more similar to their neighbors than others.

### 3.2 Multivariate case

Although the univariate models above seem quite appealing, one could be concerned about including another level in the hierarchy of the model containing that adaptive CAR distribution. Moreover, that additional layer would contain as many variables as observations in the univariate model, so this modification increases the number of variables in the model in a quite significant manner. As a consequence, data in univariate disease mapping models may be not strong enough as to make inference on vector  $\mathbf{c}$  possible. As mentioned earlier, this was already suggested by MacNab et al. (2006) and we agree with that point of view. For example, let us assume that we performed an univariate disease mapping study with an adaptive CAR distribution where the weights vector  $\mathbf{c}$  should be estimated. Let us also assume that the number of observed events for spatial unit  $i$  was abnormally higher than the corresponding number of expected cases. This would make the corresponding log-risk  $\phi_i$  take a large positive value. It seems clear that, in this case, if  $c_i$  was low, this would help  $\phi_i$  to reach that goal by allowing it to have a more independent performance, in comparison to its neighbors, and a higher variance. But, what makes the risk of this disease so strange for this spatial unit? Is it the specific particularities of the disease under study in that precise unit ( $\phi_i$ ) or the spatial unit itself that, for some structural (geographical, social, environmental, etc.) reason, is far different in general to its surrounding spatial units ( $c_i$ )? With a single observation per spatial unit the model does not have enough information to distinguish these two settings and therefore to estimate  $\mathbf{c}$  properly. In contrast, if we had several risk estimates for several diseases  $\phi_{ij}$  depending on a common weights matrix  $\mathbf{W}^*(\mathbf{c})$ , we would be able to know if the risk of the original disease in the  $i$ -th spatial unit was really different, or the differential factor was the spatial unit. In the first case, among all the log-risks of the  $i$ -th spatial unit, only that corresponding to the original disease should take a high value and  $c_i$  should not therefore take a low value since that unit does not have a different performance in general than its

neighbors. In the second, all (or most of) the log-risks for the  $i$ -th spatial unit would take extreme values and  $c_i$  should take a low value in order to accommodate that behavior. As a consequence, the use of adaptive CAR distributions in the context of multivariate studies could improve the fit of the spatial weights vector  $\mathbf{c}$  to a considerable extent. This issue will become clearer in the real case study in the next section.

In accordance to the previous paragraph, we now introduce now a multivariate model integrating adaptive CAR distributions, one per disease, with a common spatial weights matrix. This formulation allows an appropriate estimation of the vector  $\mathbf{c}$  and therefore an appropriate estimation of the weights matrix  $\mathbf{W}^*(\mathbf{c})$  corresponding to the set of diseases and region of study considered. The following formulation implements an adaptive BYM model per disease, although a similar formulation could be analogously proposed for the case of the Leroux et al. CAR distribution.

Let  $O_{ij}$  represent the observed number of cases for the  $i$ -th spatial unit and  $j$ -th disease,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ . We will assume:

$$O_{ij} \sim \text{Poisson}(E_{ij}R_{ij}),$$

where  $E_{ij}$  is the number of expected cases, and  $R_{ij}$  the relative risk for the corresponding spatial unit and disease. In accordance with the univariate model, the log-risks can be expressed as:

$$\log(R_{ij}) = \mu_j + \eta_{ij}, \tag{6}$$

where  $\mu_j$  stands for an intercept modeling the mean of the log-risks for each disease and  $\eta_{ij}$  are random effects accounting for spatial or unstructured variability for those risks. We will model the columns of  $\boldsymbol{\eta} = (\eta_{ij})$  by means of a BYM model, i.e.  $\eta_{ij} = \phi_{ij} + \theta_{ij}$ , where  $\theta_{ij}$  are independent Normal random effects and the columns of  $\boldsymbol{\phi} = (\phi_{ij})$  follow adaptive ICAR distributions, all of them depending on a common weights matrix  $\mathbf{W}^*(\mathbf{c})$  with a common weights vector  $\mathbf{c}$ , as described in the univariate case. Also, in parallel to the univariate case, the components of  $\mathbf{c}$  will be assumed to follow a  $\text{Gamma}(\alpha, \alpha)$  distribution. Different standard deviations would be considered for the columns of the  $\boldsymbol{\phi}$  and  $\boldsymbol{\theta}$  matrices in order to allow the relative risk geographical patterns to show more or less spatial dependence. The prior distribution for  $\alpha$  will be chosen similarly to the univariate case above.

Note that the model just introduced, although posed in a multivariate setting, does not take into account the hypothetical dependence between diseases that these could show. In that case, that proposal could incorporate and take advantage of that dependence. In fact, we have explored that possibility by proposing a multivariate M-model (Botella-Rocamora et al., 2015) with a common adaptive spatial weights matrix and therefore a common spatial dependence structure. We have not noticed any evident benefit of that proposal in terms of the estimation of the spatial weights matrix in comparison to the proposal above. Since our goal in this paper is focused on that estimation, we have preferred to pose the 'independent' multivariate version in order to keep its content

simpler. In any event, if the main goal was to estimate the risks  $\mathbf{R}$  for several diseases with an adaptive spatial dependence structure, the use of pure multivariate models, as the M-model mentioned, would obviously yield a significant improvement.

## 4 Application

### 4.1 Spatial weights matrices estimation from multivariate data sets

In this section, we evaluate the performance of the multivariate adaptive spatial model described in the previous section in some real scenarios. The main data set for this analysis corresponds to the observed and expected deaths in the city of Valencia (Spain), for a total of 15 different mortality causes in men for the period from 1996 to 2015. Mortality data are available for each of the 531 census tracts of Valencia, the geographical unit for this analysis. The main goal of this analysis is to estimate a suitable weights matrix for the Valencian census tracts that reflect the dependence structure of mortality causes in general over the whole city. We will use the multivariate adaptive extension developed in Section 3.2 for both the BYM and Leroux models, and the mortality data set described to estimate that matrix. Subsequently, in the next subsection, the estimated spatial structure matrix will be used in posterior univariate analyses in order to assess the improvement that its use could bring, in comparison to the traditional adjacency criterion that assumes fixed weights, equal to 1, for each adjacent pair of units.

Both BYM and Leroux et al. multivariate models were run in WinBUGS and the corresponding R code for all the analyses in this section can be found as annex material to this paper in the document `RCode.pdf`. For each model, three chains were run with 200,000 iterations, whose first 50,000 iterations were used as a burn-in period. Of these, one of every 150 iterations was saved yielding a final sample size of 3,000 iterations. Convergence was assessed by means of the Brooks-Gelman-Rubin statistic (we required this to be lower than 1.1 for each variable in the model) and the effective sample size (required to be at least 100 for each variable in the model).

We start by taking a look to the estimated weights vector  $\mathbf{c}$  for our analyses. These weights should reflect the strength of spatial dependence between each neighboring pair of spatial units. For the BYM model, the values of the spatial weights  $\mathbf{c}$  (their posterior means) range from 0.098 to 2.042, with a mean value of 1.240, while for the Leroux et al. model these values range from 0.027 to 1.886, with a mean value of 1.264. Figure 1 shows the corresponding  $c_i$  for each geographic unit for both models run. The census tracts with dark red color represent those with a lower estimated spatial weight  $c_i$ . According to the comments of Section 3, these census tracts should show a different behavior in comparison to their surrounding census tracts and, as a consequence, the model tries to separate/isolate those units. In contrast, the census tracts in yellow

are those that have been found to have a greater influence on the risk of the surrounding areas or, in other words, those most dependent on their neighbors. As shown in this figure, both adaptive proposals of the BYM and Leroux models estimate a closely similar spatial dependence structure for the region of study. The correlation between the estimated spatial weights vector  $\mathbf{c}$  for the adaptive BYM and Leroux models is 0.956, which shows the agreement of the spatial dependence structure estimated for both models.

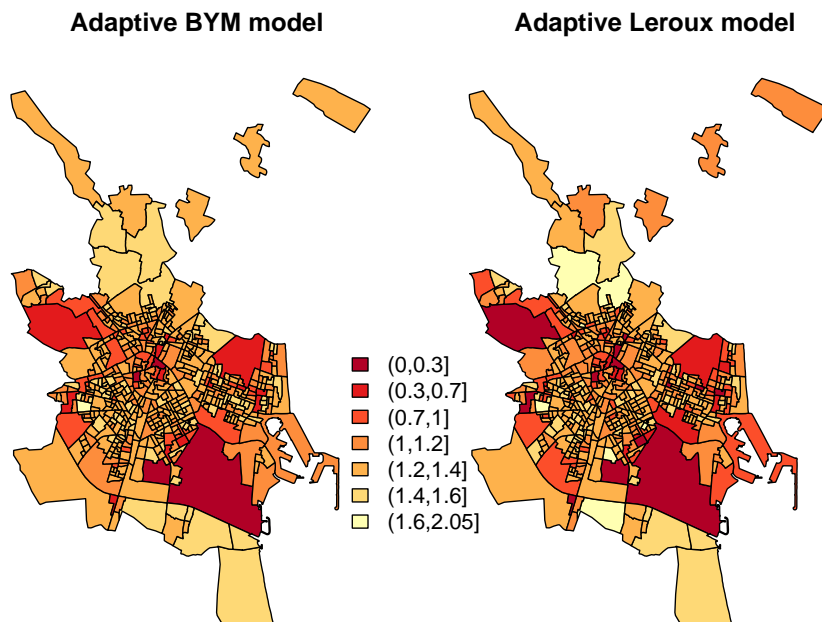


Figure 1: Estimated spatial weight  $c_i$  for each census tract of Valencia according to all 15 diseases in the data set. Each choropleth map corresponds to either BYM (left) or Leroux et al. (right) models for the log-relative risks.

We have explored the results in Figure 1 in order to interpret the spatial dependence structure estimated by the models. In particular, we have observed that the census tracts with lowest  $c_i$  values have certain peculiarities that make them special with respect to their adjacent units. On the one hand, residential homes for elderly or socially excluded people are frequently located in some of those "special" census tracts. As a consequence, these units often show higher observed deaths than expected for most of the mortality causes considered, which makes them exhibit a different behavior from those of their neighbors. Anyway, if our main goal was just to detect and model spatial units of this kind, with a different behaviour than their neighbors, the use of models accounting for discontinuities (Knorr-Held and Raßer, 2000; Denison and Holmes, 2001; Adin

et al., 2019) would be possibly a more suitable modeling choice. On the other hand, new building areas and socially marginal regions of the city also frequently show  $c_i$ s in the darkest red zones. The use of a broad set of mortality data, with 15 diseases, has allowed the models to identify those census tracts with these particularities that lead them to exhibit a very particular behavior in terms of mortality. That behavior requires an adaptation of the spatial weights matrix, otherwise their risks would be oversmoothed and estimated more similarly to their neighbors than they should. The low estimated value of their spatial weights will allow them to show the separate behavior that they require. As will be shown in the next subsection, a more flexible modeling of the risks is obtained in this manner, avoiding the excessive smoothing problems that many of the most frequent disease mapping models entail (Richardson et al., 2004; Botella-Rocamora et al., 2012). In contrast, Figure 1 shows some other regions where high spatial weights have been fitted. Note that several of these units are located in spatial units at the borders of the graph where the geometry of the graph would impose lesser spatial dependence. See for example the yellow unit in the north of the city or those in the southeast of the city, which connect some other spatial units in its south which are not completely shown at the choropleth maps. The high values of the spatial weights vector seem to be used to connect more tightly those regions of the map in the outskirts that would otherwise have an excessively independent behavior, preventing them from being isolated. Thus, the adaptive proposal run seems to change some geometric properties of the graph that could make some census tracts less connected to the rest of the graph than would be desirable.

The two tables shown in the `AdditionalResults.pdf` document reinforce also these results. The first of these tables show, for each disease, the mean absolute difference between the risks of the adjacency and adaptive models. These results are separately presented for the regions with low, medium and high weights. These results point out that the main risk differences for both models occur there where the spatial weights vector takes more extreme values, more different to 1. Thus, this is where these two models particularly differ. On the other hand, the second of these tables show the risk differences between each area and its neighbours for the regions taking low and high weights, respectively. In average, the spatial units taking low weights show higher differences in comparison to their neighbours than the adjacency based model. Thus, these low weights allow these units to have a more independent performance. In contrast, the spatial units with high weights have risk estimates more similar to their neighbours than the corresponding adjacency based estimates. Therefore, the performance of these regions is just the opposite of that of the regions taking low spatial weights.

Similar conclusions are drawn from a parallel analysis that we have made for the whole of Spain at the municipal level (see more detailed results for this analysis in the supplementary material). In this case, a multivariate analysis of 18 mortality causes has been carried out and a common spatial structure is estimated from this analysis. As a summary, 7 of the 10 municipalities with the lowest spatial weights (those which are disconnected) are municipalities in

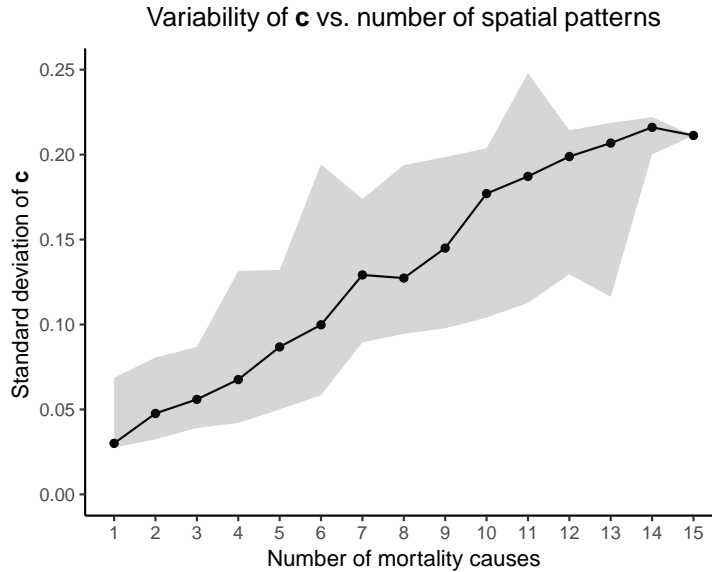


Figure 2: Variability of  $\mathbf{c}$  (standard deviation) in the Valencia city data set when estimated with the adaptive multivariate BYM model as a function of the number of diseases considered in the analysis. The black line connects the mean of the observed standard deviations of  $\mathbf{c}$  and the gray band delimits the minimum and maximum observed standard deviations of  $\mathbf{c}$  for each number of diseases.

the Costa Blanca, a Spanish region with a large foreigner community of elderly retired people from northern Europe who have moved to this Spanish region. The presence of this community has been shown to have a clear impact on the mortality of this region; see Zurriaga et al. (2008). The three remaining municipalities with low spatial weights are all provincial capitals, which correspond to municipalities with substantially higher population than their neighbors. The spatial adaptive model used seems to have been sensitive to both data artifacts, making these municipalities different to their neighbors. On the contrary, we have found that 8 out of the 10 municipalities with the highest spatial weights belong to coastal municipalities, that is, they are placed at the borders of the graph of the region of study, which seems to confirm the impression that we have drawn from the Valencia city data set.

Before concluding the study of the estimated spatial weights vector, we want to illustrate the importance of the multivariate feature of the models implemented for that estimation. Figure 2 shows, once again for the BYM model in the Valencia city data set, the variability (standard deviation) of the estimated vector  $\mathbf{c}$  (its posterior mean) as a function of the number of diseases considered. Thus, for one disease we have run 15 different models, one per disease, and we



have repeated this for another 15 (randomly chosen) pairs of diseases, 15 trios and so forth until reaching groups of 14 diseases. The black line in Figure 2 connects the mean of the observed standard deviations of  $\mathbf{c}$  for each number of diseases considered. The gray band delimits the minimum and maximum standard deviations observed for  $\mathbf{c}$  for each number of diseases. Figure 2 shows how the multivariate model proposed describes an increasing trend for the variability of  $\mathbf{c}$  as a function of the number of diseases considered. Thus, for the univariate studies,  $\mathbf{c}$  hardly learns from the data, which means that the resulting spatial weights matrix closely resembles the adjacency based weights matrix. In other words, the adaptive feature of the model has hardly any effect when a low number of diseases is considered. By way of contrast, Figure 2 shows substantial variability in  $\mathbf{c}$  when the number of diseases is higher. Thus, in summary, Figure 2 clearly points out the need to perform multivariate studies if inference is pursued for the spatial dependence structure of a region of study.

## 4.2 Use of the estimated spatial weights matrix in subsequent univariate studies

Once the spatial weights matrix of the spatial random effects has been estimated for a region, it could be used for subsequent univariate disease mapping analyses on that same region. That is, the multivariate estimates above of  $\mathbf{W}^*(\mathbf{c})$  could be used as estimates of the spatial dependence structure for later uses, instead of the traditional (although arbitrary) adjacency matrix  $\mathbf{W}$ . It is hoped that  $\mathbf{W}^*(\mathbf{c})$  would have captured the geographical structure and particularities of mortality in that region of study. In this section, we are going to assess that procedure on our data set comparing the use of the  $\mathbf{W}^*(\mathbf{c})$  matrices estimated in the previous subsection with the most traditional procedure which uses the adjacency criterion. Specifically, for all 15 diseases in our data set we have fitted univariate BYM and Leroux et al. models assuming either the spatial dependence structure estimated from the multivariate analysis above or the traditional adjacency-based weights matrix. Next, we compare the results of both analyses for each mortality cause according to the Standardized Mortality Ratios (SMR) of both alternatives and also according to the Conditional Predictive Ordinate (CPO) and the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002).

In order to make a fair comparison, avoiding the use of the data twice (once for estimating  $\mathbf{c}$  and once for estimating the SMRs with the corresponding univariate models), we have used different  $\mathbf{W}^*(\mathbf{c})$  in our comparisons. Thus, for the case of AIDS mortality, for example, we have estimated  $\mathbf{c}$  with a multivariate study of 14 diseases, all excepting AIDS, which avoids using the data twice for the univariate (preestimated) adaptive analyses. We have repeated this procedure for all 15 causes of mortality considered. Interestingly, the correlations between the spatial weights vectors  $\mathbf{c}$  for the analyses with 14 mortality causes and that with 15 mortality causes are rather high, ranging from 0.94 to 0.97 depending on the cause removed. As a consequence, we would not expect important differences if  $\mathbf{W}^*(\mathbf{c})$  had been estimated just once with all 15

diseases.

Figure 3 shows the estimated Standardized Mortality Ratios (SMR) with the BYM model for the Valencian census tracts for cirrhosis mortality (similar choropleth maps for the remaining of diseases are shown in the `AdditionalResults.pdf` file at the supplementary material of this paper). The map on the left side corresponds to the model assuming an adjacency-based relationship between spatial units, while that on the right side uses the spatial weights matrix previously estimated. The Leroux et al. model provides similar results and these are also shown in the `AdditionalResults.pdf` file.

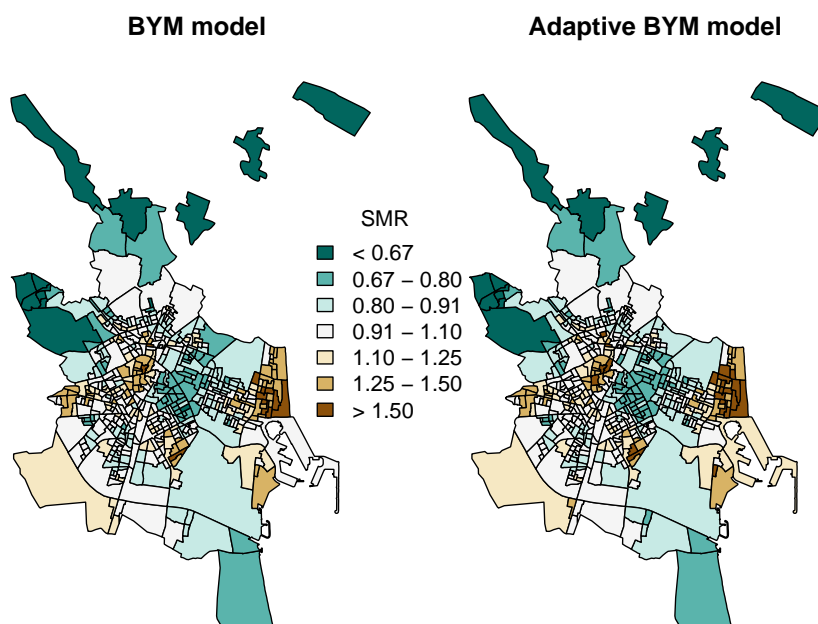


Figure 3: Standardized Mortality Ratios for Cirrhosis in Valencia estimated with the BYM model and with spatial weights matrices of either unitary weights (left) or using the values obtained from the multivariate analysis of 14 diseases (all mortality causes of study except Cirrhosis) (right).

As can be seen, both models provide risk maps with similar spatial patterns. However, the model using the adaptive spatial weights reproduces higher variability than its adjacency-based alternative allowing some of Valencia's neighborhoods to be reproduced more clearly and making it possible for some census tracts to reproduce more extreme risks. In particular, the adaptive analysis depicts more clearly some particular high risk zones scattered throughout the city which usually correspond to high deprivation areas. Thus, as previously pointed out, the adaptive weights avoid the excessive smoothing of the SMRs previously described in the literature by allowing additional flexibility wher-

ever it is required according to previous information synthesized in the previous multivariate adaptive analysis.

Afterwards, we have compared the fit of the adaptive vs the non-adaptive weights models according to the CPO and DIC criterion. Table 1 shows the DIC and CPO of the BYM and Leroux models for both spatial weights matrices considered. For each model and mortality cause we have marked the proposal providing a better fit, according to DIC and CPO, in bold. As can be observed, according to DIC (CPO), BYM and Leroux et al. models with adaptive weight matrices provide a better fit than the corresponding adjacency based model in 14 (13) and 13 (9), respectively, out of the 15 mortality causes considered. This confirms that the greater flexibility of the adaptive models really improves the SMRs estimates in comparison to the traditional adjacency-based analyses. In addition, as it can be confirmed at the annex document **AdditionalResults.pdf**, those mortality causes with a more substantial improvement in terms of DIC or CPO in general coincide with those showing a spatial pattern of higher variability (AIDS or COPD for example). Thus, an improvement is achieved mainly when there is spatial variability to be explained, otherwise the gain achieved is very modest as might seem logical.

## 5 Discussion

As described in this paper, CAR distributions are usually considered to model the spatial dependence between geographic units in disease mapping studies. Although CAR distributions are undoubtedly useful and powerful tools, the parameterization used to induce dependence by means of its structure matrix is usually arbitrary. We have introduced a procedure to estimate that spatial weights matrix according to retrospective multivariate data. As shown, our adaptive procedure makes CAR models more flexible and improves the fit of subsequent analysis adopting the estimated weights matrix, which in principle should have captured the particularities that mortality data could show in that region. Additionally, the multivariate character of our proposal has shown itself to be an indispensable tool for appropriately estimating the spatial structure of the data.

The methodology introduced could have several different uses. First, the multivariate adaptive model introduced could be used in multivariate studies with adaptive spatial structures. These models should provide more accurate risk estimates that could take advantage of the adaptive character of the spatial dependence considered. In any event, if that was the main goal of our analysis, a multivariate model, considering the dependence between mortality causes, would be much more advisable. A second use of adaptive CAR models would be the one emphasized in this paper, that is, making inference on the spatial weights matrix of a region of study. In this case, we would be more interested in the values of the weights  $\mathbf{c}$  than the own risks. As a consequence, that vector  $\mathbf{c}$ , and thus the adaptive weights matrix could be later used in subsequent enhanced spatial disease mapping studies with a non-arbitrary spatial structure based on

Table 1: DIC and CPO for the BYM and Leroux at al. models with adaptive and unweighted spatial weights matrices.

Causes		Adjacency BYM model	Adaptive BYM model	Adjacency Leroux model	Adaptive Leroux model
AIDS	DIC	1648.11	<b>1631.87</b>	1653.57	<b>1647.39</b>
	CPO	-857.98	<b>-842.16</b>	-865.67	<b>-853.49</b>
Stomach cancer	DIC	1771.57	<b>1771.43</b>	1770.46	<b>1770.30</b>
	CPO	-884.84	<b>-884.77</b>	-884.56	<b>-884.49</b>
Colorectal cancer	DIC	2354.53	<b>2354.42</b>	2354.07	<b>2353.52</b>
	CPO	<b>-1176.82</b>	-1176.95	-1176.69	<b>-1176.65</b>
Lung cancer	DIC	2861.52	<b>2857.85</b>	2872.61	<b>2872.34</b>
	CPO	-1430.85	<b>-1429.09</b>	<b>-1436.90</b>	-1438.01
Prostate cancer	DIC	2126.24	<b>2126.08</b>	2124.63	<b>2124.51</b>
	CPO	-1062.46	<b>-1062.34</b>	-1061.91	<b>-1061.83</b>
Bladder cancer	DIC	<b>2051.48</b>	2052.19	2056.07	<b>2053.80</b>
	CPO	<b>-1027.02</b>	-1027.55	-1029.16	<b>-1027.80</b>
Hematological cancer	DIC	1955.37	<b>1955.17</b>	1953.59	<b>1953.26</b>
	CPO	-977.43	<b>-977.34</b>	-976.51	<b>-976.35</b>
Mellitus diabetes	DIC	1976.97	<b>1975.01</b>	1978.35	<b>1976.38</b>
	CPO	-987.946	<b>-987.03</b>	-988.85	<b>-987.54</b>
Dementia	DIC	2335.17	<b>2333.21</b>	<b>2341.74</b>	2342.03
	CPO	-1168.47	<b>-1168.05</b>	<b>-1172.21</b>	-1172.41
Ischemic heart disease	DIC	3055.33	<b>3048.46</b>	3061.86	<b>3061.53</b>
	CPO	-1537.51	<b>-1535.37</b>	<b>-1541.93</b>	-1542.05
Ictus	DIC	2662.95	<b>2659.86</b>	2665.66	<b>2664.25</b>
	CPO	-1331.24	<b>-1329.64</b>	-1332.84	<b>-1332.18</b>
COPD	DIC	2681.81	<b>2668.87</b>	2698.39	<b>2679.29</b>
	CPO	-1348.43	<b>-1342.04</b>	-1359.55	<b>-1347.98</b>
Liver cirrhosis	DIC	2130.23	<b>2128.75</b>	<b>2139.20</b>	2141.70
	CPO	-1071.21	<b>-1070.22</b>	<b>-1075.66</b>	-1077.16
Suicides	DIC	1488.94	<b>1488.86</b>	1488.43	<b>1488.36</b>
	CPO	-744.36	<b>-744.33</b>	<b>-744.09</b>	-744.25
Traffic accidents	DIC	1506.68	<b>1506.17</b>	1506.76	<b>1506.75</b>
	CPO	-752.97	<b>-752.80</b>	<b>-753.13</b>	-753.22

previous data and knowledge.

In this sense, we find it convenient to mention a couple of limitations of the proposed methodology. Our adaptive model proposes a kind of meta-analysis of the spatial structure of several causes of death. It would be obviously convenient that these causes of death were as homogenous as possible. In an ideal situation, all of them should be cardiovascular diseases, or tumoural causes of death ... since these settings should probably share a common spatial dependence structure, as assumed by our model. Considering congenital deaths or senility, for example, as causes of death with a common spatial structure could be possibly a bit risky since that assumption would be hard to maintain in that case. Although we do not find any reason why adjacency could be a better option in this setting. Anyway, this limitation should be born in mind when using the estimated spatial weights matrix in new studies, since the new causes of death in those cases should be as related as possible to those used for estimating the spatial weights matrix.

In the same manner, as suggested by one reviewer, it would be convenient to bear in mind that adding covariates to disease mapping problems could possibly change the spatial dependence structure of the region of study. For example, if an adaptive spatial dependence structure gives a low weight to a particular spatial unit, separating it from its neighbours, this could be also reproduced by a covariate taking in this spatial unit very different values than in its neighbours. Thus, a weights matrix that could be suitable for disease mapping studies for a region of study could be not so good for ecological regression studies on that same region.

Although in principle the main uses of our model would be those mentioned in the previous paragraphs, we have also found a third practical use of the model that we did not expect. This use would be quality control of systematic problems that could be present in health data sets. Being more precise, the Valencia city mortality data used in Section 4 belongs to a large Spanish project studying mortality in large cities, the MEDEA project. All the deaths in that data set have been geocodified by using several geocoding tools, in particular the Google geocoding API and a second geocoding tool (Cartociudad) of the Spanish Geographic National Institute. These tools, as with any other geocoding tool, are not perfect and they have errors for some particular streets, groups of cases that are geocodified in the city center etc. that could distort the spatial analyses of that data base. We have found that the multivariate adaptive model on those data bases give low spatial weights to those census tracts with systematic geocoding errors since their mortality data are somewhat different from their surrounding areas. This has allowed us to distill those errors by focusing on those census tracts with low spatial weights and no potential alternative explanation (no residential homes, no socially marginal areas, no new building areas, etc.) for them. In most cases we have found that those regions contained some geocoding error. Note that the results shown in Section 4 correspond to the distilled database without geocoding errors, which have been already fixed otherwise Figure 1 would have also pointed out the census tracts with geocoding errors. This is just a single example of the many uses that adaptive CAR models could

have in practice. This work illustrates just some potential uses of adaptive CAR models, although we suspect there are many more than those that we have found. We encourage readers to keep exploring new potential uses of this approach.

## acknowledgements

The authors acknowledge the support of the research Grant PI16/01004 (co-funded with FEDER grants) of Instituto de Salud Carlos III and the predoctoral contract UGP-15-156 of FISABIO.

## References

- Adín A, Martínez-Beneito MA, Botella-Rocamora P, Goicoa T, Ugarte MD (2017) Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive and moving average models. *Stochastic and Environmental Research and Risk Assessment* 31:403–415, DOI 10.1007/s00477-016-1269-8, URL <http://dx.doi.org/10.1007/s00477-016-1269-8>
- Adin A, Lee D, Goicoa T, Ugarte MD (2019) A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Statistical Methods in Medical Research* 28(9):2595–2613, DOI 10.1177/0962280218767975, URL <https://doi.org/10.1177/0962280218767975>, pMID: 29651927, <https://doi.org/10.1177/0962280218767975>
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 36:192–236
- Besag J, Kooperberg C (1995) On conditional and intrinsic autoregressions. *Biometrika* 82(4):733–746
- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43:1–21, DOI 10.1007/BF00116466
- Best NG, Arnold R Arnold, Thomas A, Waller LA, Conlon EM (1999) Bayesian models for spatially correlated disease and exposure data. In: *Bayesian Statistics 6*, Oxford University Press
- Botella-Rocamora P, López-Quílez A, Martínez-Beneito MA (2012) Spatial moving average risk smoothing. *Statistics in Medicine* 32:2595–2612, DOI 10.1002/sim.5704

- Botella-Rocamora P, Martinez-Beneito MA, Banerjee S (2015) A unifying modeling framework for highly multivariate disease mapping. *Statistics in Medicine* 34(9):1548–1559, DOI 10.1002/sim.6423
- Brezger A, Fahrmeir L, Hennerfeind A (2007) Adaptive Gaussian Markov random fields with applications in human brain mapping. *Applied Statistics* 56(3):327–345
- Carter CK, Kohn R (1996) Markov chain monte carlo in conditionally gaussian state space models. *Biometrika* 83(3):589–601
- Congdon P (2008) A spatially adaptive conditional autoregressive prior for area health data. *Statistical Methodology* 5:552–563
- Denison DG, Holmes CC (2001) Bayesian partitioning for estimating disease risk. *Biometrics* 57:143–149
- Duncan EW, White NM, Mengersen K (2017) Spatial smoothing in bayesian models: a comparison of weights matrix specifications and their impact on inference. *International Journal of Health Geographics* 16(1):47, DOI 10.1186/s12942-017-0120-x
- Earnest A, Morgan G, Mengersen K, Louise R, Richard S, Beard J (2007) Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models. *International Journal of Health Geographics* pp 6–54
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741
- Green P, Richardson S (2002) Hidden Markov models and disease mapping. *Journal of the American Statistical Association* 97(460):1–16
- Knorr-Held L, Raßer G (2000) Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* 56(13-21):2045–2060
- Kuhnert P (2003) New methodology and comparisons for the analysis of binary data using bayesian and tree based methods
- Lang S, Fronk EM, Fahrmeir L (2002) Function estimation with locally adaptive dynamic models. *Computational statistics* 17(4):479–499
- Lawson AB (2018) *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology* (3rd edition). CRC Press
- Leroux BG, Lei X, Breslow N (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical Models in Epidemiology, the Environment, and Clinical Trials* pp 135–78

- Lu H, Carlin BP (2005) Bayesian areal Wombling for geographical boundary analysis. *Geographical Analysis* 35:265–285
- Lu H, Reilly CS, Banerjee S, Carlin BP (2007) Bayesian areal Wombling via adjacency modelling. *Environmental and Ecological Statistics* 14:433–452
- Ma H, Carlin BP (2007) Bayesian multivariate areal Wombling for multiple disease boundary analysis. *Bayesian Analysis* 2(2):281–302
- Ma H, Carlin BP, Banerjee S (2010) Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. *Biometrics* 66:355–364
- MacNab YC (2018) Some recent work on multivariate Gaussian Markov random fields. *TEST* 27(3):497–541
- MacNab YC, Kmetz A, Gustafson P, Sheps S (2006) An innovative application of bayesian disease mapping methods to patient safety research: A canadian adverse medical event study. *Statistics in Medicine* 25:3960–3980
- Martinez-Beneito MA, Botella Rocamora P (2019) *Disease Mapping: From foundations to Multidimensional Modeling*. CRC Press
- Raftery AE, Banfield JD (1991) Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics (discussion of Besag et al.). *Annals of the Institute of Statistical Mathematics* 43:32–43
- Richardson S, Thomson A, Best N, Elliot P (2004) Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives* 112(9):1016–1025
- Rue H, Held L (2005) *Gaussian Markov Random Fields: Theory & Applications*. Chapman & Hall/CRC
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583–641, DOI 10.1111/1467-9868.00353
- Zurriaga O, Vanaclocha H, Martínez-Beneito MA, Botella Rocamora P (2008) Spatio-temporal evolution of female lung cancer mortality in a region of Spain: is it worth taking migration into account? *BMC Cancer* 8(35):1, DOI 10.1186/1471-2407-8-35, URL <http://www.biomedcentral.com/1471-2407/8/35>