Spatial Moving Average Risk Smoothing

Botella-Rocamora, P.[1], López-Quílez, A.[2], Martinez-Beneito, M. A.[3,4].

[1] Universidad CEU-Cardenal Herrera.

[2] Universitat de València.

[3] Centro Superior de Investigación en Salud Pública, Valencia.

[4] CIBERESP.

3rd December 2011

Abstract

This paper introduces SMARS as a new way of carrying out disease mapping. This proposal applies the moving average ideas of time series theory to the spatial domain, indeed, it makes use of a spatial moving-average process of unknown order to define dependence on the risk of ocurrence of a disease. Correlation of the risks for different locations will be a function of m values (m being unknown) providing a rich class of correlation functions that may be reproduced by SMARS. Moreover, the distance (in terms of neighborhoods) that two units should be located to make the correlation of their risks 0 is a quantity to be fitted by the model, therefore reproducing patterns which range from spatially-independent to long-range spatially dependent.

A theoretical study of the correlation structure induced by SMARS will also be shown, illustrating the wide variety of correlation functions that this proposal is able to reproduce. Two applications of SMARS to both simulated and real datasets will also be presented. This application will show SMARS to be a competitive disease mapping model when compared with alternative proposals which have already

appeared in the literature. Finally, the application of SMARS to the study of mortality for 21 causes of death in Comunitat Valenciana will allow us to identify some qualitative differences in the patterns of those diseases.

1 Introduction

Disease mapping has undergone rapid development in recent years. Original disease mapping models have been extended to account for the temporal variability of risks in every region [1, 2], to include possible dependence among the risks of different diseases [3, 4] and to smooth other health indicators which are not necessarily risks of the presence of a disease [5, 6]. Among these extensions, and for most of the disease mapping literature in general, Gaussian Markov Random Fields [7] (GMRF) have been the most successful tools for defining dependence relations among regions. Nevertheless, a wide collection of models have been proposed with the purpose of defining new dependence structures which obtain, if possible, better smoothing properties than those of GMRF. These alternative proposals have been based on different modelling tools like, for example, tessellations of the region under study [8, 9], zero-inflated processes [10], hidden Markov models [11], mixtures of constant-risk components [12], parametric kernel smoothing [13], geostatistics [14, 15], Wombling [16, 17], Dirichlet processes [18], among others. That is, the number of tools used to describe spatial dependence in disease mapping studies is quite large but, for some reason, the use of GMRF continues to be the main way of achieving this goal. The capability of GMRF to be easily integrated into more complex models and its availability in standard Bayesian simulation software must surely be the main reason why these models have been the default tool in risk smoothing studies. Nevertheless, we are not sure if this is reason enough to justify the situation.

Disease mapping studies usually consider observed counts of a disease aggregated into a set of geographical locations. This set usually has an irregular structure, in contrast to time series studies where data are arranged in a sequential order in a unique dimension. In any case, we can consider any grid (either regular or irregular) in disease mapping

studies as a bidimensional generalization of the temporal structure in time series studies. Conversely, we can consider any time series of counts as a particular case of a disease mapping problem on a linear unidimensional grid. Therefore, the methods used in any one of these fields could be adapted to the other, providing new tools to model dependence between observations. This parallelism is going to be the starting point of this paper, where we will propose a moving average-based spatial process of unknown order for disease mapping studies.

This paper is organized as follows. Section 2 introduces some issues regarding the dependence structures induced by GMRF, motivating the ideas that we are now introducing in this paper. Section 3 contains the theoretical formulation of the Spatial Moving Average Risk Smoothing model (SMARS from now on). Section 4 makes an analysis of the dependence structure defined by SMARS and compares it with other smoothing proposals appearing in the literature. Section 5 gives the results of the application of the model to various datasets and finally Section 6 concludes with a brief discussion, including lines of future work.

2 Motivation

To illustrate the performance of the dependence structure defined by some GMRF we are going to consider several linear lattices of different lengths, as would be the case with any time series. For these regions, we will consider that any unit in the lattice has just two neighbours (the previous and posterior units), excepting the first and last observations which have just one neighbour (the second and penultimate observations, respectively).

Intrinsic Conditional AutoRegressive distributions (ICAR) are certainly the most common GMRF in disease mapping applications. These distributions, when all neighbours are equally weighted (a rather usual assumption), can be formulated as:

$$x_i|\boldsymbol{x}_{-i} \sim N\left(n_i^{-1}\sum_{\{j:j\sim i\}} x_j, n_i^{-1}\sigma^2\right) \quad i=1,...,n$$

where $\boldsymbol{x}_{-i}$ denotes the set $\{x_1,...,x_{i-1},x_{i+1},...,x_n\}$, n denotes the number of units in the

lattice of study, the set $\{j : j \sim i\}$ contains all the neighbouring locations j of unit i, and n_i denotes the number of neighbours for the i-th unit on the lattice. Therefore, the expected value at every location is just the mean of the observations for its neighbouring regions on the lattice. If the lattice of study were linear, the ICAR distribution could be shown to be equivalent to a first order Random Walk [19] or, following the ARIMA terminology from time series, a First Order Integrated Process. Therefore, we can find a strong relationship between these two tools, the Integrated Processes of time series studies and the ICAR distributions of disease mapping modelling.

If we consider an ICAR distribution on a linear lattice, the correlation of the observation at the most central location with any other place follows the relationship appearing in the upper part of Figure 1. In the upper-left part of that figure we can see how that correlation varies for a linear lattice with 51 observations, and in the upper-right side we can see that same relation for a lattice with 501 observations. As can be appreciated, both plots show the same pattern as the length of the lattice increases. Indeed, when that length increases the number of neighbourhoods that we have to move across to make the correlation between observations 0 also increases by a similar ratio, and the shape of the correlation function does not change. Moreover, this correlation function does not change if we modify the unique parameter of the ICAR distributions, the precision parameter. Therefore, the number of neighbours that we have to move across to get rid of correlation between two locations is not a modifiable parameter in ICAR distributions; however, it comes fully determined, at least on linear lattices, by the length of the region of study. As a consequence, the dependence structure induced by this distribution will depend only on the shape and length of the lattice under study. Moreover, smoothed risks on large lattices could be oversmoothed as risks will necessarily be dependent up to quite long distances. Moreover, these plots also suggest that correlations between the most distant units for ICAR distributions are negative, that is, this distribution imposes opposing risk behaviours for distant places instead of being independent, which would look much more reasonable. All these features of ICAR distributions do not look particularly attractive and they may have unknown effects on the smoothed risk estimates, mainly when the

disease studied is rarer, as in that case the information provided by data is scarcer.

[PUT FIGURE 1 JUST HERE]

However, ICAR distributions are usually accompanied by a heterogeneous random effect in disease mapping studies, as proposed by Besag et al. [20], in order to account for risk factors or features exclusively affecting individual geographical units, whose scope does not reach neighboring geographical locations. Nevertheless, the inclusion of this random effect does not alleviate most of the features outlined in the previous paragraph. As can be appreciated in the lower-left part of Figure 1, the correlation function (for the lattice of length 51) when adding the spatially independent term changes, flattening the correlation decrease as a function of distance. Moreover, that decrease gets flatter when the independent random effect has larger variability. Therefore, although the correlation function is more flexible with this change, it continues to be somewhat rigid in some aspects that could be interesting to modify in order to describe different collections of geographical patterns better.

Proper Conditional AutoRegressive distributions (CAR-proper) are the second GMRF of widespread use in disease mapping studies. These spatial modelling tools can be defined as a set of conditional distributions in the following way:

$$x_i|\boldsymbol{x}_{-i} \sim N\left(\rho \cdot \left(n_i^{-1} \sum_{\{j:j\sim i\}} x_j\right), n_i^{-1}\sigma^2\right) \quad i = 1, ..., n$$

therefore, the expected value of every observation in a CAR-proper distribution is given by a regression on the mean value of the distribution in their surrounding regions. This idea is closely related to that followed by First Order Auto-regressive time series; therefore, a clear parallelism can be established between these two modelling tools. In contrast to ICAR, CAR-proper distributions have a parameter (ρ) to fit the correlation structure to the pattern followed by the data. As can be appreciated in the lower-right part of Figure 1, this parameter allows us to control the speed of decrease in the correlation between observations as a function of the distance between them. Therefore, this distribution allows us to model patterns with very different ranges of dependence, that is, this

distribution is adaptive in this sense. Moreover, for the CAR-proper distribution, the correlation tends to zero as distance increases; therefore, the CAR-proper distribution solves the most controversial issues in the ICAR distribution.

Despite all these appealing features of the CAR-proper distribution, this is not as popular as could be hoped. Indeed, several authors have already pointed out some drawbacks to this distribution [21, 22, 23]. To those drawbacks we would like to add the following: the decrease in correlation as a function of the distance in the lower-right part of Figure 1 is a parametric function of ρ, very close to an exponential decrease depending on the value of ρ. Although this is a significant improvement compared with the rigid correlation function of the ICAR distribution, this may not be enough as any pattern not responding to this correlation structure will not be appropriately described with this distribution: that is, the parametric shape of the correlation could continue to be too rigid to explain some geographical patterns.

As already pointed out, a parallelism can be formulated for integrated processes in time and its equivalent in space (ICAR). In a similar way, auto-regressive time series can be viewed as the temporal version of the CAR-proper distribution. Nevertheless, moving average processes of time series do not have a clear equivalent in the disease mapping literature. Best et al. [24] perform a kernel-weighted smoothing that in some way makes a local mean for estimating the risk at every location. Nevertheless, they use a parametric (Gaussian) kernel to weight the information about surrounding locations, and this is not a common practice in moving average time series modelling; therefore, that work cannot be understood as the spatial moving average process that we are seeking. Finally, although some work has been done towards a spatial moving average model (and even spatial ARMA models) [25], this has always been always done assuming that the original data is Gaussian, an assumption rarely appropriate in the disease mapping context. Therefore, an effort should be made to make this methodology useful in this field.

The goal of this paper is to make a new proposal for disease mapping following the ideas of the moving average modelling of time series. In this proposal, we would like to avoid any parametric relation for determining the weights of neighbouring units in order

to be able to draw a wide collection of spatial patterns. Moreover, we will consider the range of dependence of the spatial pattern as a quantity to be fitted by the model, in order to describe patterns as either dependent for only quite close regions or patterns whose distant regions will also be dependent. This way we will try to solve the aforementioned problems of the spatial patterns arising from the ICAR and CAR-proper distributions.

3 The SMARS model

Before to formulate the SMARS model it is convenient to introduce some notation criteria and define the neighbourhood relationships that we will use in the rest of the paper. Regarding neighbourhood criteria, we will consider 0-th order neighbour of region i as its own region i; its 1-st order neighbours will be all those regions adjacent to region i (or if preferred, any other criteria based, for example, on the distance between regions) and its 0-th order neighbours; the 2-nd order neighbours will be all those regions adjacent to the 1-st order neighbours and its own 1-st order neighbours, etcetera. This way, if any region is n-th order neighbour of region i it will also be n'-th order neighbour for any $n' > n$, and for any n the set of n-th order neighbours will form a disk centred on region i and radius defined by n. This neighbourhood relationship is not the frequently-used criteria forming rings around any unit in the region of study, but we will see later that it has several advantages that justify its use in our proposal.

If regions i and j are k-th order regions, we will denote that relationship as $i \sim_k j$. We will denote the number of units in the region of study as n, the number of observed cases in unit i as O_i and the number of expected cases on that same unit as E_i.

As usual in most disease mapping models, we will consider the number of observed cases for every unit to follow a Poisson distribution:

$$O_i \sim Po(E_i \cdot exp(\phi_i)) \;\; i = 1, ..., n$$

where $\boldsymbol{\phi}$ stands for the vector of log-relative risks. We will model those log-relative risks in the second layer of the hierarchical structure. To carry this out, we will consider a

Gaussian latent risk process $\{\theta_i : \ i =, ..., n\}$ for every one of the units in the region of study. We will define the log-relative risks as the sum of an intercept term (μ), modeling the mean value of the log-risks in the region of study, plus a second term for every region proportional to a weighted sum of the latent effects in those same regions and their neighbours. This latter term incorporates the moving average structure into the relative risks. To be more precise, we will assume that the risk in every region depends on its own latent effect, plus the latent effects of the first order neighbours, plus the latent effects of the second order neighbours, etcetera, up to the neighbours of order m. We will call this parameter the *order* of the moving average process and we will assume it to be unknown. We will weight the latent effects of the neighbours of differing orders by means of the vector of positive weights $\boldsymbol{\omega} = (\omega_0, ..., \omega_m)$, also to be estimated by the model, which weights the contribution of the furthest regions compared to those closest. In summary, our proposal to model the log-relative risks in region i can be expressed as:

$$\phi_i = \mu + \lambda_i^{-1} \cdot \left(\omega_0 \cdot \theta_i + \omega_1 \cdot (\sum_{j \sim_1 i} \theta_j) + ... + \omega_m \cdot (\sum_{j \sim_m i} \theta_j) \right) . \tag{1}$$

In the previous expression, if we had $m = 0$, the risks at every unit would only depend on the latent value of that precise unit; therefore, the SMARS model would reproduce a fully heterogeneous pattern. On the contrary, for $m = 1$, the risks would also depend on the latent effects of neighbouring regions, and for higher values of m latent values of farther places would also influence the risk at any location. Therefore, depending on the value of m, SMARS may describe either independent or spatially dependent processes (either of long or short range). In the former section, we saw that ICAR-based models were not able to do this as in that case the range of the spatial patterns was fully determined by the length or shape of the lattice considered. Moreover, the values of the vector $\boldsymbol{\omega}$ determine the influence of farther neighbours (of an order lower than m) compared with the influence of closer neighbours. Furthermore, as we will mention below, the values of $\boldsymbol{\omega}$ do not follow any parametric relationship as a function of distance; therefore, the dependence structures defined by SMARS will be more flexible than those based on the CAR-proper distribution, as the latter cannot balance the importance of farther neighbours compared with those

being closer.

As the moving average process in (1) already induces spatial dependence, we will consider $\boldsymbol{\theta}$ to have independent values. That is; we will assume:

$$\boldsymbol{\theta} \sim N_n(\mathbf{0}, \sigma^2 \cdot \boldsymbol{I}_n),$$

therefore, $\boldsymbol{\theta}$ defines the values of the underlying risk at every location that $\boldsymbol{\omega}$, by means of the moving average process, propagates to the surrounding regions. That is, a high (respectively low) value of θ_i will induce an increase (decrease) in the risk for region i and for its neighbouring regions, depending on the values of the vector $\boldsymbol{\omega}$. Therefore, we can consider $\boldsymbol{\theta}$ as an unblurred distribution of risk factors which have a local influence on the distribution of the disease in nearby units and not only at the exact location of risk sources.

The spatial dependence structure defined by (1) entails some secondary effects like, for example, an edge effect, which it would be convenient to control. This is where the λ_i term in (1) plays a role. In the SMARS model, the variance of the log-relative risk for the i-th unit of study can be derived as follows:

$$\begin{aligned}
&Var(\phi_i|\mu, \boldsymbol{\omega}) = \\
&= Var\left(\lambda_i^{-1} \cdot \left(\omega_0 \cdot \theta_i + \omega_1 \cdot (\sum_{j \sim_1 i} \theta_j) + ... + \omega_m \cdot (\sum_{j \sim_m i} \theta_j)\right)\right) = \\
&= \lambda_i^{-2} \cdot \left(\omega_0^2 Var(\theta_i) + \omega_1^2 \cdot (\sum_{s \sim_1 i} Var(\theta_j)) + ... + \omega_m^2 \cdot (\sum_{s \sim_m i} Var(\theta_j))\right) = \\
&= \lambda_i^{-2} \cdot \left(\sigma^2 \cdot (\omega_0^2 + \omega_1^2 n_1(i) + ... + \omega_m^2 n_m(i))\right),
\end{aligned}$$

where $n_1(i), ..., n_m(i)$ are, respectively, the number of neighbours of first, second, ..., m-th order of region i. Therefore, the variance of the log-relative risk in every region is a function of the number of neighbours, favouring the occurrence of edge effects. To correct this behaviour we propose to define:

$$\lambda_i = \left(\omega_0^2 + \omega_1^2 n_1(i) + ... + \omega_m^2 n_m(i)\right)^{1/2}, \tag{2}$$

and this way all the components of $\boldsymbol{\phi}$ (also the latent effects) will have exactly the same variance (σ^2) regardless of their locations in the region of study.

For the rest of this section, we are going to propose and justify the prior distributions for all the parameters in the SMARS model. Our intention is to restrict the weights $\boldsymbol{\omega}$ so that they take positive values as, in that case, the contribution of closer units in the moving average process will be necessarily higher than that of farther locations, which seems reasonable. This is a consequence of the way in which we have defined neighbour relationships forming disks instead of rings. In any case, we will provide more details about this issue in the next section. A side effect of the inclusion of the normalizing terms λ_i in (1) is that the scale of $\boldsymbol{\omega}$ cannot be identified as a consequence of the dependence of λ_i on $\boldsymbol{\omega}$. That is, the weights of the moving average process for the i-th unit $\lambda_i^{-1}(\omega) \cdot \boldsymbol{\omega}$ (in this case we include $\lambda_i^{-1}(\omega)$ to emphasize its dependence on $\boldsymbol{\omega}$) are equal for $\boldsymbol{\omega} = \boldsymbol{w}$ and $\boldsymbol{\omega} = \alpha\boldsymbol{w}$. In more detail, for any vector $\boldsymbol{w}$ and positive scalar α we have:

$$\lambda_i^{-1}(\alpha\boldsymbol{w}) \cdot (\alpha\boldsymbol{w}) = \left((\alpha w_0)^2 + (\alpha w_1)^2 n_1(i) + ... + (\alpha w_m)^2 n_m(i)\right)^{-1/2} \cdot (\alpha\boldsymbol{w}) =$$

$$= \left(w_0^2 + w_1^2 n_1(i) + ... + w_m^2 n_m(i)\right)^{-1/2} \cdot \alpha^{-1} \cdot (\alpha\boldsymbol{w}) = \lambda_i^{-1}(\boldsymbol{w}) \cdot \boldsymbol{w}$$

As a consequence, if no further restriction is imposed, the scale of $\boldsymbol{\omega}$ may not be identified. Therefore, we have considered $\boldsymbol{\omega}$ to follow a Dirichlet prior distribution of parameter $\mathbf{1}_{n+1}$, that is a uniform distribution on the range of values of $\boldsymbol{\omega}$. The Dirichlet distribution implicitly imposes $\sum_{i=0}^{m} \omega_i = 1$ and this way we avoid the aforementioned identifiability issues on the scale of $\boldsymbol{\omega}$. In summary, we consider as a prior distribution of $\boldsymbol{\omega}$:

$$(\omega_0, \omega_1, ..., \omega_m)|m \sim \text{Dir}(\mathbf{1}_{m+1}).$$

This choice has an additional advantage of considerable importance for the final results of the model. Indeed, SMARS is one of those models where '*the number of things that you do not know is one of the things that you do not know*' [26] as the length of $\boldsymbol{\omega}$ depends on the parameter m, which is also unknown. Therefore, we can view SMARS as an encompassing model containing a collection of models with a fixed number of parameters and we will be interested in selecting one of them or quantifying their probabilities. This is clearly a model selection problem. A clear distinction has been established in the literature between common and non-common parameters for the models considered

[27]. It has been described that, for model selection purposes, it is possible to propose arbitrarily vague (or even improper) distributions for the common parameters without conditioning the posterior probability of the models considered. On the contrary, for the non-common parameters ($\boldsymbol{\omega}$ in our case) the choice of arbitrarily vague prior distributions has a significant effect on the posterior distribution of the models, making them depend on the degree of vagueness of the priors used [28]. Therefore, the choice of the prior distribution for $\boldsymbol{\omega}$ is a tricky issue for the final performance of the model. The choice of a Dirichlet prior distribution for $\boldsymbol{\omega}$ means that the probability space for this vector is finite; indeed, it is a simplex on an $m+1$-dimensional space, where the uniform prior distribution is an obvious and reasonable choice for avoiding the use of arbitrarily vague proposals.

The order of the spatial moving average process m may take any value on the set $\{0, 1, ...\}$, although for practical reasons we will assume an upper bound (M) for this variable. This bound is motivated by the fact that for any m bigger than any specific m' all units will be m-order neighbours; therefore, due to the finite nature of the lattice of study, it makes no sense to consider neighbours of indefinitely large orders. A uniform prior on the range of admissible values for m could seem in principle to be a reasonable proposal, although that choice would not enable the value of m to be identified. To illustrate this, let us assume that the real value for $(m, \boldsymbol{\omega})$ is $(1, (0.5, 0.5))$. The prior probability of that value is:

$$p(m, \boldsymbol{\omega}) = p(\boldsymbol{\omega}|m) \cdot p(m) = \frac{\Gamma(2)}{\Gamma(1) \cdot \Gamma(1)} \cdot \frac{1}{M+1} = \frac{1}{1 \cdot 1} \cdot \frac{1}{M+1} = \frac{1}{M+1},$$

while the prior probability of, for example, $(m, \boldsymbol{\omega}) = (4, (0.5, 0.5, 0, 0, 0))$ would be:

$$p(m, \boldsymbol{\omega}) = p(\boldsymbol{\omega}|m) \cdot p(m) = \frac{\Gamma(4)}{\Gamma(1)^5} \cdot \frac{1}{M+1} = \frac{4!}{1^5} \cdot \frac{1}{M+1} = \frac{4!}{M+1}.$$

The likelihood function will yield identical values for both pairs $(m, \boldsymbol{\omega})$ considered (as they yield identical values for $\boldsymbol{\phi}$). Therefore, if the prior distribution does not penalize large values of m, the model will tend to fit large values for this variable regardless of its true value, making it unidentifiable. As a consequence, we propose the prior distribution

of m to be:

$$p(m) \propto \frac{1}{m!}\ , \quad m = 0, ..., M.$$

as this way, the prior distributions of m and $\boldsymbol{\omega}$ compensate for one another and all the values of the pair $(m, \boldsymbol{\omega})$ will have exactly the same prior probability.

For the rest of the common parameters included in the model we have proposed vague prior distributions. Thus, we have proposed an improper flat prior distribution on the whole real line for the mean value of the log-relative risks (μ). We have also proposed an improper flat prior distribution on the positive real line for the standard deviation of the latent effect (σ) as suggested, for example, in Gelman (2005) [29].

4 An insight into the dependence structure

Having introduced our proposal in the previous section, we are now going to explore the dependence structure just defined in some detail. We will also compare the previous model and its dependence structure with other proposals in the literature, pointing out their similarities and differences.

First, we are going to express the previous model in matricial notation which will be helpful for the rest of this section. We will denote the k-th order neighbourhood matrix as $\boldsymbol{H}_k$, that is, for this matrix the cell (i, j) will be 1 if and only if locations i and j are k-th order neighbourhood, otherwise that cell will be 0. We will define $\boldsymbol{H}_0$ as the $n \times n$ identity matrix because, as previously defined, each unit is the only 0-th order neighbour of itself. In that case, we can put expression (1) as:

$$\begin{aligned} \boldsymbol{\phi} &= \mu \mathbf{1}_n + diag(\boldsymbol{\lambda}^{-1}) \left(\omega_0 \boldsymbol{H}_0 \boldsymbol{\theta} + \omega_1 \boldsymbol{H}_1 \boldsymbol{\theta} + ... + \omega_m \boldsymbol{H}_m \boldsymbol{\theta}\right) = \\ &= \mu \mathbf{1}_n + diag(\boldsymbol{\lambda}^{-1}) \left(\omega_0 \boldsymbol{H}_0 + \omega_1 \boldsymbol{H}_1 + ... + \omega_m \boldsymbol{H}_m\right) \boldsymbol{\theta}\ , \end{aligned}$$

and if we denote $(\omega_0 \boldsymbol{H}_0 + \omega_1 \boldsymbol{H}_1 + ... + \omega_m \boldsymbol{H}_m)$ as $\boldsymbol{\Lambda}$ we can express:

$$\boldsymbol{\phi} = \mu \mathbf{1}_n + diag(\boldsymbol{\lambda}^{-1}) \boldsymbol{\Lambda} \boldsymbol{\theta}\ ,$$

where, if the distance between units i and j is $d(<= m)$, $\Lambda_{ij} = \sum_{k=d}^{m} \omega_k$, and if $d > m$ then $\Lambda_{ij} = 0$. Therefore, we can consider the log-relative risks as a scaled version (to avoid edge effects) of the transformed latent factor $\boldsymbol{\Lambda\theta}$.

In the preceding expression, $\boldsymbol{\lambda}$ is a vector with components given by (2). These values have a direct relation with $\boldsymbol{\Lambda}$ as (it is straightforward to show that) λ_i is just the Euclidean norm of the i-th row of $\boldsymbol{\Lambda}$. Therefore, we may state this last expression as:

$$\boldsymbol{\phi} = \mu \mathbf{1}_n + (diag(\boldsymbol{\Lambda\Lambda'})^{-1/2}\boldsymbol{\Lambda})\boldsymbol{\theta}$$

In summary, we can express:

$$\boldsymbol{\phi} = \mu \mathbf{1}_n + \hat{\boldsymbol{\Lambda}}\boldsymbol{\theta} \tag{3}$$

where:

$$\hat{\boldsymbol{\Lambda}} = (diag(\boldsymbol{\Lambda\Lambda'})^{-1/2}\boldsymbol{\Lambda})$$

and

$$\boldsymbol{\Lambda} = (\omega_0 \boldsymbol{H}_0 + \omega_1 \boldsymbol{H}_1 + ... + \omega_m \boldsymbol{H}_m)\,.$$

Note that both $\boldsymbol{\Lambda}$ and $\hat{\boldsymbol{\Lambda}}$ are a function of m and $\boldsymbol{\omega}$, so they are stochastic matrices, both in their sparse structure and the values of the non-zero cells. Moreover, the matrix $\hat{\boldsymbol{\Lambda}}$ inducing the correlation structure on $\boldsymbol{\phi}$ is row-standardized in the sense that each one of their rows has 1 as Euclidean norm.

We are now in a position to have a deeper insight into the dependence structure defined by SMARS.

4.1 The dependence structure of the SMARS proposal

Taking into account that $\boldsymbol{\theta}$ is distributed as $N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ and expression (3), the prior distribution of $\boldsymbol{\phi}$ can be derived as:

$$\boldsymbol{\phi}|\mu, \hat{\boldsymbol{\Lambda}} \sim \mathrm{N}_n(\mu \mathbf{1}_n, \sigma^2 \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Lambda}}') = \mathrm{N}_n(\mu \mathbf{1}_n, \sigma^2 diag(\boldsymbol{\Lambda\Lambda}')^{-1/2} \boldsymbol{\Lambda\Lambda}' diag(\boldsymbol{\Lambda\Lambda}')^{-1/2})\,.$$

Therefore, the covariance of any two log-relative risks in SMARS can be expressed as:

$$\mathrm{Cov}(\phi_i, \phi_j|\mu, \hat{\boldsymbol{\Lambda}}) = \sigma^2 \frac{\boldsymbol{\Lambda}_{i\cdot} \cdot (\boldsymbol{\Lambda'})_{\cdot j}}{\sqrt{\boldsymbol{\Lambda}_{i\cdot} \cdot (\boldsymbol{\Lambda'})_{\cdot i}} \sqrt{\boldsymbol{\Lambda}_{j\cdot} \cdot (\boldsymbol{\Lambda'})_{\cdot j}}} = \sigma^2 \frac{\boldsymbol{\Lambda}_{i\cdot} \cdot \boldsymbol{\Lambda}_{j\cdot}}{\sqrt{\boldsymbol{\Lambda}_{i\cdot} \cdot \boldsymbol{\Lambda}_{i\cdot}} \sqrt{\boldsymbol{\Lambda}_{j\cdot} \cdot \boldsymbol{\Lambda}_{j\cdot}}}\,,$$

where in the preceding expression $\boldsymbol{\Lambda}_{i\cdot}$ and $\boldsymbol{\Lambda}_{\cdot j}$ stand for the i-th row and j-th column of $\boldsymbol{\Lambda}$, respectively. As $\boldsymbol{\Lambda}_{i\cdot} = \sum_{k=0}^{m} \omega_k (\boldsymbol{H}_k)_{i\cdot}$, it results that:

$$\begin{aligned}\boldsymbol{\Lambda}_{i\cdot} \cdot \boldsymbol{\Lambda}_{j\cdot} = \\ = \left(\sum_{k=0}^{m} \omega_k (\boldsymbol{H}_k)_{i\cdot}\right)\left(\sum_{l=0}^{m} \omega_l (\boldsymbol{H}_l)_{j\cdot}\right) = \sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l \left((\boldsymbol{H}_k)_{i\cdot} \cdot (\boldsymbol{H}_l)_{j\cdot}\right) .\end{aligned}$$

In the previous expression, $((\boldsymbol{H}_k)_{i\cdot} \cdot (\boldsymbol{H}_l)_{j\cdot})$ is exactly the number of units in the region of study which are k-order neighbours of unit i and l-order neighbours of unit j. We will denote that quantity as $n_{[k,l]}(i,j)$. As with our neighbourhood definition, for any $l > k$, any k-order neighbour is also a l-order neighbour; we have that $n_{[k,l]}(i,i)$ will be exactly $n_{min(k,l)}(i)$ for any k and l, that is the number of $min(k,l)$-order neighbours in unit i. Therefore, the former expression of the covariance can be expressed as:

$$\begin{aligned}&\mathrm{Cov}(\phi_i, \phi_j|\mu, \boldsymbol{\omega}) = \\ &= \sigma^2 \frac{\sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l n_{[k,ll]}(i,j)}{\sqrt{\sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l n_{min(k,l)}(i)}\sqrt{\sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l n_{min(k,l)}(j)}} .\end{aligned} \tag{4}$$

Moreover, it can also be deduced that:

$$Var(\phi_i|\mu, \boldsymbol{\omega}) = Cov(\phi_i, \phi_i|\mu, \boldsymbol{\omega}) = \sigma^2 \frac{\sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l n_{[k,l]}(i,i)}{\sum_{k=0}^{m}\sum_{l=0}^{m} \omega_k\omega_l n_{min(k,l)}(i)} = \sigma^2 .$$

Therefore, the correlation between any two log-relative risks i and j coincides with expression (4) without the σ^2 term. Therefore, those correlations will depend on two factors, the lattice's own structure and the weights of the different orders of neighbourhood. On the one hand, the correlations will be larger for those regions with more paths connecting them, although units with more neighbours will lead in general to lower correlations with surrounding units. On the other hand, in contrast to ICAR or CAR-proper distributions, we will have several parameters available (as many as considered necessary by the model) in order to adapt correlation values to the features of any dataset considered. Moreover, every correlation value between the log-relative risks fitted by SMARS will never be negative, as each one of the terms either in the numerator or the denominator of expression (4) is non-negative. That expression also shows that, for any two regions whose distance

is larger than $2m$ neighbours, their correlation will be 0, as in that case all the terms in the numerator of (4) will be 0.

As an example of the correlation structure of SMARS, we have considered the same linear lattice ($n = 51$) as in Section 2, and we have plotted the correlation of the central location of the lattice with the rest of the locations for different values of the pair $(m, \boldsymbol{\omega})$. These correlation functions are shown in Figure 2. On the left-hand side of that figure, we can see the effect of changing the value of m for vectors $\boldsymbol{\omega}$ with equal values for all their components. It can be appreciated how farther regions with positive correlations can be found when the value of m is increased. Conversely, on the right-hand side of Figure 2, we can see the effect of changing the vector $\boldsymbol{\omega}$, keeping the value of m constant. In this case, the shape of the decrease is modified by the values of $\boldsymbol{\omega}$, illustrating the flexibility of SMARS for describing a wide range of relationships between distances and correlations.

[PUT FIGURE 2 JUST HERE]

4.2 SMARS modelling and kernel smoothing

Another proposal also used in several disease mapping applications is the Kernel Convolution approach [13]. This approach, in order to define spatially dependent log-relative risks, considers an underlying spatially independent process $\boldsymbol{\theta}$ convolved with a parametric kernel function k . If the underlying risk is defined on a continuous domain $\mathcal{D}$ it yields:

$$\phi_i = \mu + \int_{\mathcal{D}} k(s_i, u)\theta(u)du \ ,$$

on the other hand, if $\boldsymbol{\theta}$ is defined in a discrete domain it yields:

$$\phi_i = \mu + \sum_{j=1}^{n} k(s_i, s_j)\theta_j \Rightarrow \boldsymbol{\phi} = \mu \mathbf{1}_n + \boldsymbol{\Delta} \cdot \boldsymbol{\theta} \ , \tag{5}$$

where s_i and s_j are the locations of units i and j respectively. Row i of matrix $\boldsymbol{\Delta}$ contains the distance-based weights needed to combine the values of the random effects $\boldsymbol{\theta}$ to define ϕ_i. The kernel function k is usually assumed to be a non-negative, symmetric and integrating one; therefore, the values of $\boldsymbol{\Delta}$ will also be non-negative, an isotropic

function of distance and in some way row-standardized. Moreover, the definition and properties of the kernel function will determine the covariance structure of the smoothed process [30].

The right-hand side of (5) coincides with the matricial expression of the SMARS structure of the log-relative risks in (3) for $\boldsymbol{\Delta} = \hat{\boldsymbol{\Lambda}}$, where the rows of $\hat{\boldsymbol{\Lambda}}$ are proportional (in order to normalize those rows) to:

$$\Lambda_{ij} = \begin{cases} (\sum_{l=d}^{m} \omega_l) & if\ d := distance(s_i, s_j) < m \\ 0 & if\ d := distance(s_i, s_j) > m \end{cases}$$

It is obvious that the values of $\hat{\boldsymbol{\Lambda}}$ are non-negative by construction, being a direct isotropic function of the distance between regions and its rows are standardized (as previously noted $\|\hat{\boldsymbol{\Lambda}}_{i\cdot}\|_2 = 1$). Therefore, we can consider the SMARS model as a kernel smoothing proposal on a discrete domain, with a kernel function (weighting the contribution of the geographical units as a function of their distance) given by the vector $(\sum_{l=d}^{m} \omega_l)_{d=0}^{m}$. We will call this vector the kernel function of the model.

As a direct consequence of the disk-shaped neighbourhood relationship used for the SMARS model formulation, the vector $\boldsymbol{\Lambda}_{i\cdot}$ puts more weight in the smoothing process of ϕ_i on those regions that are closer to region i. That is, if d is the distance between regions i and j, the latent factor θ_j will be multiplied by a value proportional to $(\sum_{l=d}^{m} \omega_l)$. Therefore, closer regions will be weighted more (and as a consequence will be more influential) in the calculation of ϕ_i. The vector

$$(\sum_{l=0}^{m} \omega_l, \sum_{l=1}^{m} \omega_l, ..., \sum_{l=m}^{m} \omega_l) = (1, \sum_{l=1}^{m} \omega_l, ..., \omega_m) \tag{6}$$

defines the unscaled transformation of the latent factor $\boldsymbol{\theta}$ as a function of the distance between regions; therefore, their components draw the shape of the (unscaled) kernel function weighting the values of the latent factor. As just pointed out, this kernel will be a decreasing function of the distance between regions, which seems reasonable. Moreover, it is also interesting that the kernel function defined by SMARS does not follow any parametric decrease (Gaussian, exponencial, triangular, ...) as is usual in kernel smoothing

methods. Therefore, in this sense, SMARS is a more flexible approach to kernel smoothing than other kernel smoothing proposals with parametric kernel functions.

The (unscaled) kernel function (6) fully controls the correlation structure of the log-relative risks. Therefore, the study and comparison of these functions for different geographical patterns, corresponding to different diseases, is another interesting tool that SMARS makes available to quantify differences between those patterns studied. Therefore, taking into account to that function we can distinguish between fully heterogenous patterns ($(m, \boldsymbol{\omega}) = (0, (1))$), patterns weighting equally all those regions up to a fixed distance (for example $(m, \boldsymbol{\omega}) = (3, (0, 0, 0, 1))$), patterns with a short-range spatial dependence (for example $(m, \boldsymbol{\omega}) = (1, (0.3, 0.7))$), or long-range dependence (for example $(m, \boldsymbol{\omega}) = (7, (0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1))$). Figure 3 shows some examples of kernels defining remarkably different spatial patterns. These could be labelled as heterogenous ($m = 0$), short range ($m = 2$) and large range patterns ($m = 4$). For the $m = 2$ scenario two different kernels are shown, one of them weighting their neighbours more than the other and, therefore, providing qualitatively distinct patterns. In our opinion, that figure yields a useful epidemiological tool to explore the dependence structure of the patterns studied and, on occasions, to determine significant differences between patterns in terms of that structure. In the applications section we will give further details on this tool.

[PUT FIGURE 3 JUST HERE]

4.3 SMARS modelling and the SAR distribution

SAR Gaussian distributions are another tool for inducing spatial dependence in disease mapping studies [31]. These distributions correspond to a vector of variables following:

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{D}((\boldsymbol{I} - \boldsymbol{B})^{-1})'),$$

where $\boldsymbol{B}$ is a matrix containing the neighbourhood structure of the region under study and $\boldsymbol{D}$ is a diagonal matrix. From (3) it is straightforward to show that we can express $\boldsymbol{\phi}$ as:

$$\boldsymbol{\phi} \sim N_n(\mu \mathbf{1}_n, \hat{\boldsymbol{\Lambda}}(\sigma^2 \boldsymbol{I}_n)\hat{\boldsymbol{\Lambda}}').$$

Therefore, we can consider SMARS as defining a distribution from the SAR family, for $\boldsymbol{D} = \sigma^2 \boldsymbol{I}_n$ and $\boldsymbol{B} = \boldsymbol{I}_n - \hat{\boldsymbol{\Lambda}}^{-1}$. Therefore, in contrast to SAR models, the matrix ($\boldsymbol{I}$-$\boldsymbol{B}$) is not sparse but is its inverse. That is, with SMARS we are directly modelling the variance-covariance matrix of the patterns instead of its precision matrix, as is usually done for Gaussian Markov Random Fields. This has already been pointed out as an attractive modelling property for the development of new disease mapping proposals [22, 23].

A second difference between SMARS and SAR distributions is that for the first, the neighbourhood matrix is not fixed but dependent on the pair $(m, \boldsymbol{\omega})$. Nor is the sparse structure of that matrix fixed as it depends on the value of m, which is also stochastic for the model. Thus, that matrix is expected to be flexible enough to fit the dependence structure of data.

5 SMARS performance assessment

To make inferences on this model for the datasets below, an MCMC algorithm has been developed. The corresponding MCMC procedure has been coded in R [32], making use of the package *spam* [33] for storage and computation with sparse matrices. R routines to run the model are available at the url `http://maphysco.ceuuch.es/~pbotella/SMARS.pdf`.

The parameters sampled in the MCMC process are: $(\mu, \sigma, \boldsymbol{\theta}, \boldsymbol{\omega}, m)$. μ has been sampled by Gibbs sampling. Analytical integration of μ on the posterior distribution of σ is possible, therefore, once integrated, μ slice sampling has been used to generate new values for σ in the MCMC. The whole vector $\boldsymbol{\theta}$ has been sampled in a single movement of the MCMC instead of sampling their components one by one in a sequential order. The Langevin-Hastings method has been used to sample from this vector, making smarter proposals for new values of $\boldsymbol{\theta}$. Vector $\boldsymbol{\omega}$ has been sampled with the Metropolis-Hastings algorithm. Finally, m has been sampled by means of a split-merge type Reversible Jump algorithm, as any change in this value entails modification of the length of vector $\boldsymbol{\omega}$. This step proposes either splitting the last component of $\boldsymbol{\omega}$ into two separate components (decomposing ω_m into new values ω_m^* and ω_{m+1}^*) or merging these two components into

a unique value making $\omega_m^* = \omega_m + \omega_{m+1}$ and $\omega_{m+1}^* = 0$. Sampling from both $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ requires fixing a band-width parameter for the respective Random Walk Metropolis-Hastings algorithms. An adaptative tuning of these parameters has been implemented during the burn-in phase of the algorithm in order to carry out a semi-automatic choice of them, improving the convergence of both vectors.

For every dataset, three independent chains were run with 5.000 burn-in iterations and 20.000 additional draws, and of these 1 in every 20 values was recorded, yielding a final posterior sample of 3.000 iterations for every parameter. More details on the whole code and the MCMC algorithms can be found in Botella-Rocamora (2010) [34].

5.1 Comparison with other Bayesian spatial methods

Several Bayesian methods have been developed in recent years proposing different ways to carry out spatial risk smoothing. To put all these proposals in order, Best et al. (2005) [35] perform a comparison of all them, taking into account several different aspects. As a first step, in order to assess the hypothetical benefits of our proposal, we have run SMARS on the same dataset used to carry out the above-mentioned model comparison (the dataset was provided by the authors of the paper). This artificial dataset consists of 5 independent draws (replicates) of a common underlying spatial process. The comparison between models for these replicates was performed based on the Deviance Information Criterion (DIC) [36]. The models to be compared in this study are the following:

- EXP: Multivariate normal distribution for log-risks with covariance between units being an exponencial function of their distance [37].
- BYM: Convolution of spatially structured and heterogenous random effects [20].
- MIX: Mixture model for risks with a spatially dependent allocation of units to components of the mixture by means of a Potts model [11].
- KHR: Modelling based on the partition of the region under study by means of tesellations [8].

- GMA: Parametric kernel smoothing of a Gamma underlying process [13].
- SMARS: The proposal introduced in this paper.

Table 1 shows the results obtained by SMARS in the former comparison (results partially appearing in Best et al. (2005) [35]). As can be appreciated, SMARS yields the lowest DIC in 2 out of 5 datasets. Moreover, its performance may be qualified as quite robust as when it does not yield the lowest DIC the difference with the best model is usually rather moderate (in DIC terms). In contrast, the mixture model (MIX) gets the lowest on 3 out of 5 occasions, but its performance is not as robust as SMARS. These results seem to suggest that SMARS is, at least, a competitive proposal for carrying out spatial risk smoothing and that going beyond the usual convolution process in Besag et al. (1991) [20] can bring considerable advantages from a goodness-of-fit point of view.

It is also interesting to note the coherence of the parameter estimates in SMARS for all five datasets considered. The posterior mode of m has been 1 for one of the datasets, 2 in three occasions and 3 for the last dataset. The posterior mean of this parameter for all five datasets varies from 1.6 to 2.4 (8 being the maximum value that m can take in these datasets), therefore suggesting a common value of this parameter around 2. Figure 4 shows the estimation of the kernel function for all five datasets. As can be noticed, all five kernel functions are quite similar with hardly any contribution of the underlying effect of those regions located four or more units away. Taking into account that all five datasets were realizations of a common underlying process, these results seem quite coherent.

[PUT FIGURE 4 JUST HERE]

5.2 The Mortality Atlas of Valencia Community

We have also run SMARS in a real context, on the data from the 'Mortality Atlas of Valencia Community, 1991-2000' [38] previously published following the convolution model of Besag et al. (1991) [20]. In that Atlas 22 causes of death were studied for men and 21 for women, although for simplicity we will only present our results for men (the results

for women were not substantially different). Mortality was aggregated by municipality, an administrative division consisting of 540 units for the Valencia Community (one of the 17 regions making up Spain, with around 4 million inhabitants in the period of study).

We find it particularly interesting to explore the estimates of the kernel function for all these causes of mortality. Figure 5 shows those kernels classified as a function of the posterior mode of m. Those causes in the upper-left plot (Arteriosclerosis, Colorectal Cancer, Cirrhosis, Hipertension, Other Heart Diseases, Prostate Cancer and Rectum Cancer) have 0 as the posterior mode for m, therefore the geographical patterns for these diseases have almost no spatial dependence. The upper-right plot contains those causes with the posterior mode of m being 1 (Colon Cancer, Leukemia, Lymphatic Cancer, Bladder Cancer), these patterns do have spatial dependence but of a very short range, with only quite close units showing dependent estimates of risks. The lower-left plot contains those causes with the posterior mode of m being 2 (Mouth Cancer, Diabetes, Liver Cancer, Ischaemic Heart Disease, Larynx Cancer and Pancreatic Cancer), these diseases having a moderate range of dependence. Finally, the lower-right plot contains those diseases with a long range of dependence, $m = 3$ or 4 (Cerebrovascular Disease, Respiratory Infections, Lung Cancer, All Cancers, Chronic Obstructive Pulmonary Disease).

[PUT FIGURE 5 JUST HERE]

It is interesting to note that diseases related to the respiratory system (or maybe tobacco-related diseases), in general, have geographical patterns which are qualitatively distinct to the other causes of mortality. Indeed, those diseases usually have long-range dependence patterns, reflecting a geographically smooth distribution of those risk factors related with those diseases; in contrast, for example, with those diseases related to the digestive system as Colon or Rectum Cancer whose risk factors, according to our results, are distributed heterogenously throughout the region of study. These kinds of conclusions cannot be drawn, for example, from the convolution model, as in that case results may only indicate a more substantial contribution of the heterogeneous or the spatial effect. No specific inference about the range of dependence can be made in that case, and in our

opinion this is an interesting result from an epidemiological point of view.

Correlations between risk estimates from the convolution and SMARS models for the different causes of mortality range from 0.76 to 0.99. Nevertheless, the values of those correlations depend on the estimated value of m. Indeed, the mean correlation for both models for those causes in the upper-left plot of Figure (5) is 0.90, for those diseases in the upper-right plot 0.85, for those in the lower-left plot 0.91 and those in the lower-right plot 0.97. Therefore, when SMARS reproduces long-range dependence, its results closely reproduce those derived from the BYM model. This is evidence of the ability of the Intrinsic CAR distribution to reproduce these kinds of patterns, as already pointed out in Section 2. This correlation decreases as m gets lower values, except for the first plot ($m = 0$), as that reproduces independence between geographical units and the convolution model includes a specific random effect to model that particular situation. Nonetheless, this result shows the difficulties of the convolution model when trying to reproduce patterns with short-range spatial dependence. In that case, SMARS is a preferable option as its flexibility also makes it possible to model dependence in that setting.

6 Conclusions

This paper has introduced SMARS as a new proposal for carrying out disease mapping. SMARS applies the moving average ideas of time series theory to the spatial domain. The ability of SMARS to estimate the order of the spatial moving average process leads to a flexible class of models able to fit a wide collection of patterns that may respond to very different correlation structures. Moreover, spatial dependency will depend on $m + 1$ parameters (for m unknown and only determined by the data) providing a flexible tool able to fit quite different correlation structures. In this sense SMARS can fit patterns of either heterogeneous, short or long range of dependence and, as outlined in this paper, this is not such a trivial task for many other proposals. Finally, the dependence structure defined by means of SMARS does not depend on the size of the region of study.

The results suggest that SMARS is a competitive alternative to most of the disease mapping models already proposed in the literature. Moreover, SMARS-based inference has several added-value products from an epidemiological point of view. For example, the ability to quantify the range of dependence (m) of the patterns of study is important as it sheds light on the geographical distribution of risk factors and the geographical distance of their effect. On the other hand, kernel functions could be an interesting summary of the degree of similarity between diseases. For example, quite different kernel functions for two diseases could suggest the existence of different risk factors (of different geographical diffusion) determining these diseases.

Acknowledgements

Paloma Botella-Rocamora and Antonio López-Quílez would like to thank the Ministerio de Educación y Ciencia for financial support (jointly financed with European Regional Development Fund) via the research Grant MTM2010-19528 and of the Generalitat Valenciana via the research Grant ACOMP11/218. We would also like to thank Nicky Best by providing us the simulated datasets used in section 5.1.

References

[1] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 2000; **19**:2555–2567.

[2] Martínez-Beneito MA, López-Quílez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* 2008; **27**:2874–2889.

[3] Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 2003; **4**(1):11–25.

[4] Jin X, Banerjee S, Carlin BP. Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2007; **69**(5):817–838.

[5] Assunção RM, Potter JE, Cavenaghi SM. A Bayesian space varying parameter applied to estimating fertility schedules. *Statistics in Medicine* 2002; **21**:2057–2075.

[6] Li Y, Ryan L. Modeling spatial survival data using semiparametric frailty models. *Biometrics* Jun 2002; **58**(2):287–97.

[7] Rue H, Held L. *Gaussian Markov Random Fields: Theory & Applications*. Chapman & Hall/CRC, 2005.

[8] Knorr-Held L, Rasser G. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* 2000; **56**(13-21):2045–2060.

[9] Denison D, Holmes C. Bayesian partitioning for estimating disease risk. *Biometrics* 2001; **57**:143–149.

[10] Ugarte MD, Ibañez B, Militino AF. Testing for poisson zero inflation in disease mapping. *Biometrical Journal* 2004; **46**:526–539.

[11] Green P, Richardson S. Hidden Markov models and disease mapping. *Journal of the American Statistical Association* 12 2002; **97**(460):1–16.

[12] Fernández C, Green P. Modelling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):805–826.

[13] Best N, Ickstadt K, Wolpert RL, Briggs DJ. Combining models of health and exposure data: The SAVIAH Study. *Spatial Epidemiology-Methods and Applications*, Elliott P, Wakefield JC, Best N, Briggs DJ (eds.). Oxford University Press, 2000; 393–414.

[14] Kelsall JE, Wakefield J. Modeling spatial variation in disease risk: A geostatistical approach. *Journal of the American Statistical Association* 2002; **97**:692–701.

[15] Goovaerts P, Gebreab S. How does Poisson Kriging compare to the popular BYM model for mapping disease risks? *International Journal of H* 2008; **7**:6.

[16] Lu H, Carlin BP. Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis* 2005; **35**:265–285.

[17] Lu H, Reilly CS, Banerjee S, Carlin BP. Bayesian areal wombling via adjacency modelling. *Environmental and ecological statistics* 2007; **14**:433–452.

[18] Kottas A, Duan JA, Gelfand AE. Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal* 2007; **49**:1–14.

[19] Besag J, Green P, Higdon D, Mengersen K. Bayesian computation and stochastic systems. *Statistical Science* 1995; **10**(1):3–41.

[20] Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathemathics* 1991; **43**:1–21.

[21] Banerjee S, Carlin BP, Gelfand AE. *Hierarchical modelling and analysis for spatial data*. Chapman & Hall/CRC, 2003.

[22] Wall MM. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* 2004; **121**:311–324.

[23] Assunçao RM, Krainski E. Neighborhood dependence in Bayesian spatial models. *Biometrical Journal* 2009; **51**:851–869.

[24] Best NG, Ickstadt K, Wolpert RL, Briggs DJ. Spatial poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association* 2000; **95**:1076–1088.

[25] Huang J. The autorregressive moving average model for spatial analysis. *Australian Journal of Statistics* 1984; **26**:169–178.

[26] Richardson S, Green PJ. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1997; **59**(4):731–792.

[27] Jeffreys H. *Theory of Probability, third edition*. Oxford University Press, 1961.

[28] Berger JO, Pericchi LR. *Model Selection*, chap. Objective Bayesian Methods for Model Selection: Introduction and Comparison. Institute of Mathematical Statistics, 2001; 135–207.

[29] Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2005; **1**(3):515–533.

[30] Higdon D. A primer on space-time modelling from a bayesian perspective. *Spatio-temporal Modelling*, Finkenstädt B, Held L, Isham V (eds.). Chapman & Hall/CRC, 2007.

[31] Cressie NA. *Statistics for spatial data. Revised edition*. John Wiley & Sons, 1993.

[32] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2009. ISBN 3-900051-07-0. URL: http://www.R-project.com.

[33] Furrer R, Sain SR. spam: A sparse R package with emphasis on mcmc methods for gaussian markov random fields. *Journal of Statistical Software* 2010; **36**(1):1–25.

[34] Botella-Rocamora P. Spatio-temporal smoothing in disease mapping (in spanish). PhD Thesis, Universitat de València 2010. URL `http://maphysco.ceuuch.es/~pbotella/TESIS_PalomaBotella.pdf`, available at `http://maphysco.ceuuch.es/~pbotella/TESIS_PalomaBotella.pdf`.

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

[36] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**:583–641.

[37] Diggle PJ, Tawn JA, Moyeed RA. Model based geostatistics (with discussion). *Applied Statistics* 1998; **47**(3):299–350.

[38] Martínez-Beneito MA, López Quílez A, Amador A, Melchor I, Botella Rocamora P, Abellán C, Abellán J, Verdejo F, Zurriaga O, Vanaclocha V, *et al.*. *Atlas de mortalidad de la Comunidad Valenciana, 1991-2000*. Generalitat Valenciana, 2005.

Dataset	EXP	BYM	MIX	KHR	GMA	SMARS
1	+31.2	+12.4	+14.7	+9.3	+16.9	**193.1**
2	+18.7	+1.2	**205.5**	+8.5	+18.1	+5.5
3	+33.7	+3.1	**192.9**	+1.8	+40.0	+2.2
4	+39.4	+14.7	+18.7	+5.5	+38.8	**168.2**
5	+22.4	+12.1	**168.8**	+17.8	+22.2	+12.0

Table 1: DIC for the models considered in the study. The lowest DIC for every dataset appears in bold and the remaining values are the DIC differences between the corresponding model and the one with lowest DIC for that dataset. Results partially appear in [35].

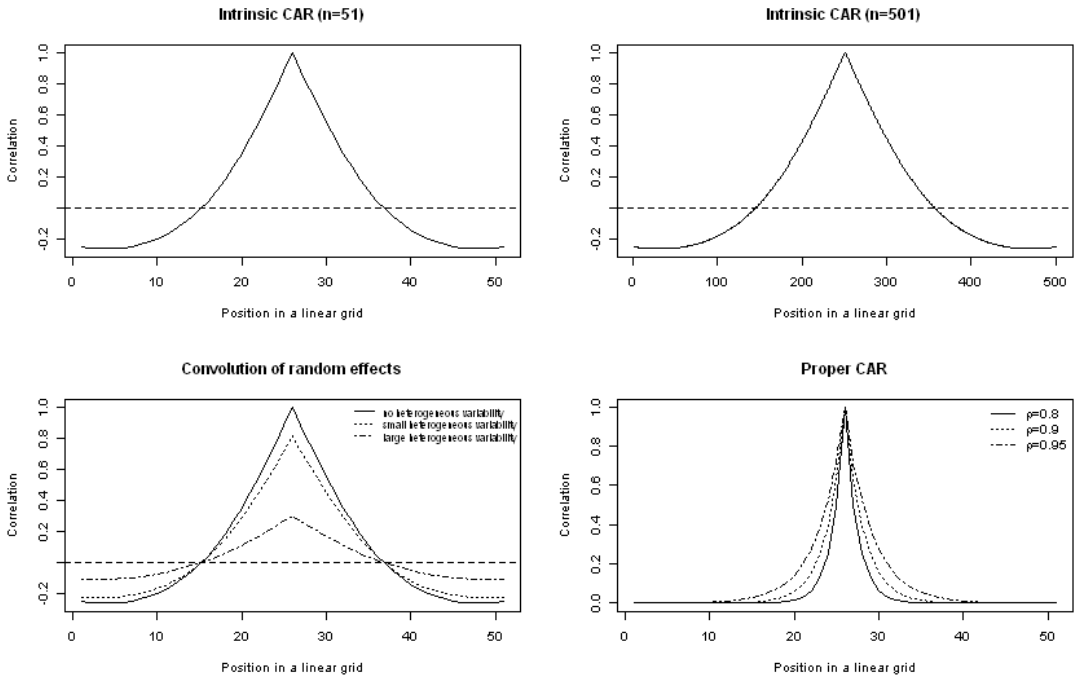

Figure 1: Correlation between the central observation and any other position on a linear lattice. Upper-left plot: correlation for an ICAR distribution on a linear grid of 51 units. Upper-right plot: correlation for an ICAR distribution on a linear grid of 501 units. Lower-left plot: correlation for a convolution model with different contributions of the heterogeneous random effect. Lower-right plot: correlation for CAR-proper distributions with different values of their parameters.

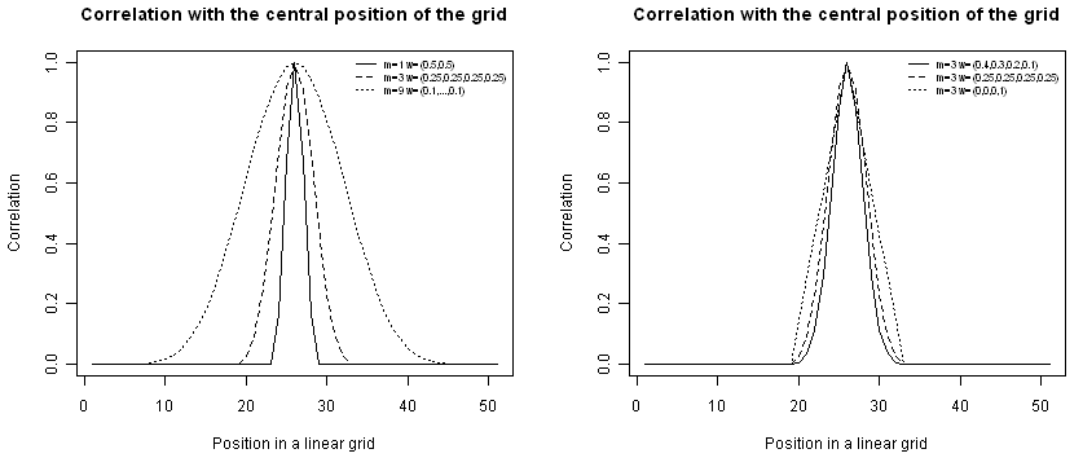

Figure 2: Correlation between the central observation and any other position for the SMARS model in a linear lattice. Left plot: Correlations for different values of m and components of $\boldsymbol{\omega}$ taking the same value. Right plot: Correlations for $m = 3$ and different values of the vector $\boldsymbol{\omega}$.

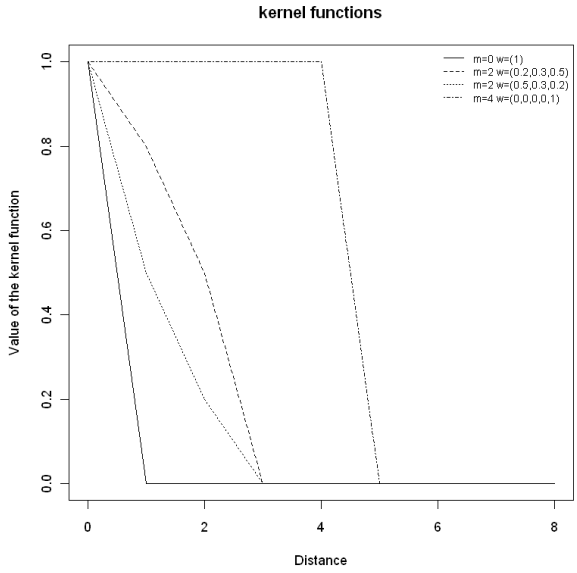

Figure 3: Effect of the parameters m and $\boldsymbol{\omega}$ on the kernel function of the SMARS model.

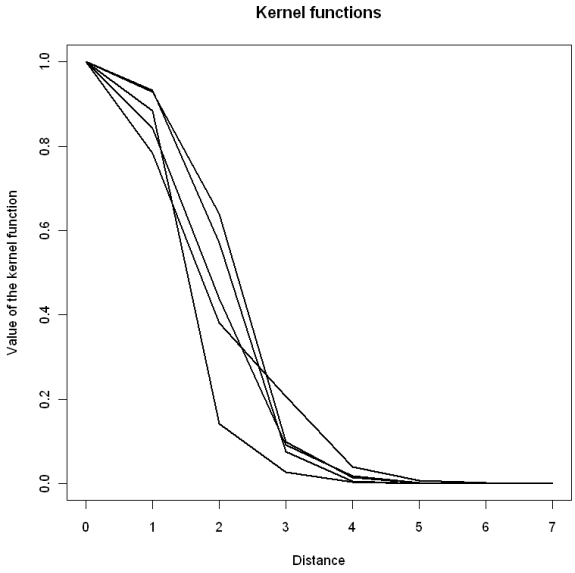

Figure 4: Effect of the parameters m and $\boldsymbol{\omega}$ on the kernel function of the SMARS model for all five datasets in Best et al. (2005)

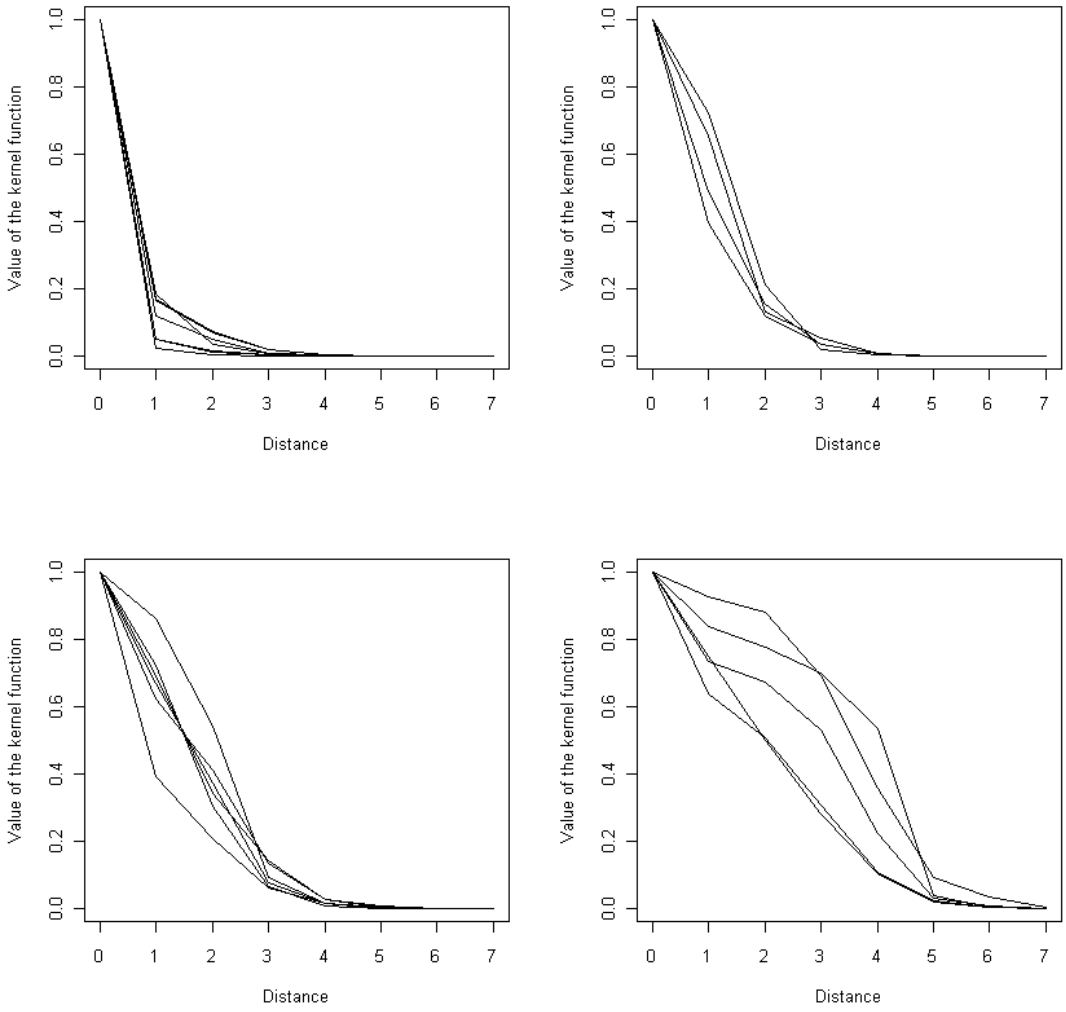

Figure 5: Estimated kernel functions for the diseases in the Mortality Atlas. Plots correspond to those diseases with posterior mode for m being 0, 1, 2 and more, respectively.