

# On the convenience of heteroscedasticity in highly multivariate disease mapping\*

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

Botella-Rocamora, Paloma  
*Epidemiology, Health Surveillance and Environmental Health Subdirectorate*  
*Conselleria de Sanidad, Valencia, Spain.*

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

## Abstract

Highly multivariate disease mapping has recently been proposed as an enhancement of traditional multivariate studies, making it possible to perform the joint analysis of a large number of diseases. This line of research has an important potential since it integrates the information of many diseases into a single model yielding richer and more accurate risk maps. In this paper we show how some of the proposals already put forward in this area display some particular problems when applied to small regions of study. Specifically, the homoscedasticity of these proposals may produce evident misfits and distorted risk maps. In this paper we propose two new models to deal with the variance-adaptivity problem in multivariate disease mapping studies and give some theoretical insights on their interpretation.

## 1 Introduction

The analysis of geographical variations in rates of diseases has a long tradition in epidemiology and statistics. This area of research, known as disease mapping, has generated substantial interest from a methodological point of view. In the beginning, disease mapping studies focused mainly on the modeling of a

---

\*This is a preprint version of the paper: Corpas-Burgos, F; Botella-Rocamora, P; Martinez-Beneito, MA. On the convenience of heteroscedasticity in highly multivariate disease mapping. *Test*, 2019, 28, 1229-1250.

single disease. However, there may be several diseases with common shared risk factors. Recently, multivariate disease mapping has received considerable attention by researchers interested in the simultaneous joint spatial modeling of several diseases (MacNab, 2016; Macnab, 2016; Martinez-Beneito et al., 2017). Multivariate disease mapping models attempt to estimate the risk of a disease in specific locations by using its spatial dependence as well as the geographical distribution of the risks for other related diseases. By so doing, a greater amount of information is used in the estimation of the risks than in univariate models, which allows more precise estimates to be obtained.

Multivariate disease mapping models are often computationally slow, and therefore most of the existing literature is restricted to multivariate modeling of two or three diseases at the most. Recently, Martinez-Beneito (2013) developed a general framework for multivariate disease mapping capable of reproducing many of the Bayesian multivariate disease mapping models previously proposed in the literature. Subsequently, Botella-Rocamora et al. (2015) extended the previous work by developing a simpler and computationally more convenient form that could handle a considerably large number (tens) of diseases. A second important advantage of this proposal is that it can be implemented in regular Bayesian simulation packages such as WinBUGS (Lunn et al., 2000).

In this work, we present an application of the methodology proposed in Botella-Rocamora et al. (2015) for the spatial modeling of several diseases in the cities of Alicante, Castellón and Valencia, which belong to the Valencian region, one of the 17 administrative regions that Spain is divided into. After observing the results obtained, some limitations of the previous methodology are evidenced when it is applied to smaller cities, as is the case of Castellón. For this reason, we propose an enhancement, variance adaptivity, of Botella-Rocamora et al.'s methodology, which allows the problems evidenced to be solved and thereby improving multivariate risk estimates.

This paper is organized as follows. Section 2 describes the modeling proposal in Botella-Rocamora et al. (2015) for multivariate disease mapping. Section 3 shows an application of the previous methodology to real data in the Spanish cities of Alicante, Castellón and Valencia. In Section 4 we propose a modification of the previous model that makes it possible to solve some of the problems found in the estimation of the risk maps for the city of Castellón. Section 5 presents and compares the results obtained with the new modeling proposal. Finally, Section 6 contains some conclusions about the models and the results obtained in the previous sections.

## 2 The $M$ -model for multivariate disease mapping

A general statistical framework for the multivariate disease mapping problem can be described as follows. Let  $O_{ij}$  and  $E_{ij}$  denote, respectively, the number of observed and expected cases for the  $i$ -th geographical unit of study and the

$j$ -th disease. The data likelihood assumes that

$$O_{ij} \sim \text{Poisson}(E_{ij}RR_{ij}) \quad i = 1, \dots, I, \quad j = 1, \dots, J$$

where  $RR_{ij}$  is the relative risk for the  $i$ -th geographical unit and  $j$ -th disease, and is modeled as  $\log(RR_{ij}) = \mu_j + \theta_{ij}$ . The term  $\mu_j$  is just an intercept for the  $j$ -th disease and  $\Theta = \{\theta_{ij} : i = 1, \dots, I; j = 1, \dots, J\}$  is a collection of random effects whose joint distribution specifies how dependence is defined within and between diseases. Specifically, dependence among the columns of  $\Theta$  induces dependence between diseases and, similarly, dependence among their rows induces spatial dependence within diseases (geographical units).

The original modeling proposal in Botella-Rocamora et al. (2015) induces spatial multivariate dependence by setting

$$\Theta = \Phi \mathbf{M} \tag{1}$$

where  $\Phi$  is an  $I \times K$  matrix of random effects with independently distributed columns that typically follow some spatially correlated distribution, such as a proper CAR (Conditional Auto-Regressive) distribution. Nevertheless, several columns with different dependence structures could also be used to reproduce more complex spatial dependence structures, such as Besag, York and Mollié's model (Besag et al., 1991), BYM henceforth. Those spatial distributions induce dependence between spatial units and therefore between rows of  $\Theta$ . Additionally,  $\mathbf{M}$  is a  $K \times J$  random matrix which induces dependence between the different columns in  $\Theta$ , that is, between the different diseases considered in the analysis. Usually  $K = J$ , although they could be different, such as for the multivariate formulation of the BYM model, where two random effects are included per disease and therefore  $K = 2J$ . The variance parameter of the random effects in the  $\Phi$  columns is usually fixed to some value so that  $\mathbf{M}$  is responsible for controlling the variability of  $\Theta$ . On the other hand, as proposed by Botella-Rocamora et al., the cells of  $\mathbf{M}$  are independently defined as  $M_{ij} \sim N(0, \sigma^2) \quad i = 1, \dots, K, \quad j = 1, \dots, J$ , where  $\sigma$  could be either a fixed (typically large) value, and therefore the  $M_{ij}$ s would follow vague independent prior distributions, or an additional variable to be estimated in the model. In the first case, we call the corresponding modeling *fixed effects M-modeling*, since  $\mathbf{M}$  cells would be modeled in that manner and, alternatively, we call the second case *random effects M-modeling*, once again because of the modeling of the cells carried out in  $\mathbf{M}$ .

To conclude this brief introduction to Botella-Rocamora et al.'s proposal, we believe it is also wise to mention a theoretical property of this model that will be used later in this work. Thus, as shown in the original paper, assigning  $N(0, \sigma^2)$  prior distributions to the entries in  $\mathbf{M}$  yields a  $Wishart(K, \sigma^2 \mathbf{I}_J)$  prior distribution for the covariance matrix between diseases  $\Sigma_b$  when all spatial models share the same spatial distribution, which can be computed as simply  $\mathbf{M}'\mathbf{M}$ . Hence, the independent modeling of the cells of  $\mathbf{M}$  entails a prior mean for  $\Sigma_b$  proportional to an identity matrix or, alternatively, it assumes prior independence in the columns of  $\Theta$ .

## 3 A motivating analysis

### 3.1 Multivariate mortality study in Castellón

The multivariate proposal put forward by Botella-Rocamora et al. (2015) has been implemented to study the geographical distribution of mortality in the cities of Alicante, Castellón and Valencia. In this section, we present some results obtained in the city of Castellón, which was composed of 95 census tracts (the geographical unit for this analysis) and had around 170,000 inhabitants in 2016. Parallel results for the analyses performed in Alicante and Valencia, composed of 215 and 553 census tracts respectively, are included as supplementary material to this paper due to lack of space (see supplementary file `Motivating_Ali_Val.pdf`). We consider the multivariate joint spatial modeling of 20 different causes of mortality and both *fixed* and *random effects M-models* for all three cities separately. In order to evaluate the benefits of multivariate modeling, we compare the results obtained with Botella-Rocamora et al.'s model with underlying BYM spatial patterns against those obtained with independent BYM patterns for each disease.

All models were run in `WinBUGS` and the R code for calling each of them can be found as annex material in `RCode.pdf`. Three chains were run for each model with 30,000 iterations, the first 5,000 of which were discarded as burn-in period. Of these, one out of every 75 iterations was saved, thereby yielding a final sample size of 1,002 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). Convergence was assessed with the `R2WinBUGS` package of R.

Figure 1 shows the results obtained with univariate BYM models (upper row), fixed effects **M**-modeling (middle row), and random effects **M**-modeling (lower row) for 3 out of the 20 causes of death under study in Castellón: AIDS, Cerebrovascular disease, and Suicides in men. Results shown for the fixed effects **M**-model assume improper  $M_{ij} \propto 1$  distributions, that is, we implicitly assume  $\sigma$  to be set to  $\infty$  in this case. Nevertheless, we have also run the same model with  $\sigma$  set to high fixed values, such as 100 or 1000, obtaining results that are barely distinguishable. Green colors correspond to census tracts with estimated low risks (Smoothed Standardized Mortality Ratio (sSMRs)  $< 0.67 = (1.5)^{-1}$  for darker greens), while brown colors correspond to units of high risk (sSMRs  $> 1.5$  for darker browns). Light-colored units denote milder deviations from the overall risk for the city.

As can be appreciated, markedly different risk maps are obtained with the multivariate fixed effects **M**-model, as compared to the univariate BYM models. Although the risk maps for AIDS for both models do not present such marked differences (a map with great variability is obtained for both models), in the case of Cerebrovascular disease and Suicides, quite distinct risk maps are obtained. While univariate modeling generally provides maps with low variabil-

ity, fixed effects  $\mathbf{M}$ -modeling provides maps with great variability, with hardly any smoothing, which resemble the corresponding maps of unsmoothed SMRs (not shown in the paper). This performance of the fixed effects  $\mathbf{M}$ -modeling in Castellón has also been observed for most of the diseases in the study. Interestingly, this lack of smoothing is noticed, but to a much lesser extent, in the results drawn from Alicante and Valencia (see the previously mentioned `Motivating_Ali_Val.pdf` document).

The lower row in Figure 1 shows the results for these same three diseases for the random effects  $\mathbf{M}$ -model. As can be seen, in this case, there are no major differences between the risk maps of Cerebrovascular disease and Suicides for the independent BYM models and the random effects  $\mathbf{M}$ -model. Both models show low variability and similar geographical patterns. However, the two risk maps obtained for AIDS mortality are dramatically different. The univariate model points out several census tracts with extreme risk in some specific locations in the city of Castellón that are known to be quite deprived. In contrast, a much flatter map (more similar in this sense to those of Cerebrovascular disease and Suicides) is obtained with the random effects  $\mathbf{M}$ -modeling. On this map, no census tract shows high risk, as is also the case for Cerebrovascular disease and Suicides. In general, we observed that random effects  $\mathbf{M}$ -modeling estimates in Castellón yield flat risk maps for all 20 diseases studied, which in a few cases, such as AIDS, are very different from those estimated with univariate modeling. Strikingly, this performance was only seen in Castellón, but not (or not so evident) in Alicante or Valencia.

In the next subsection we will attempt to explain why these strange results are obtained for the fixed and random effects  $\mathbf{M}$ -models in the city of Castellón.

### 3.2 A statistical interpretation of the results in the motivating analysis

First, we find it interesting to emphasize that far milder differences were found between the fixed and random effects risk patterns in Alicante and Valencia. For these two cities, both multivariate models take advantage of the additional information provided by the set of diseases considered, depicting more detailed spatial patterns in general than their univariate alternatives. This suggests that the results found for Castellón could be due to the smaller size of this city, where the prior structure that the  $\mathbf{M}$ -model induces could be more influential than in Alicante and Valencia. Thus, the prior covariance structure of the  $\mathbf{M}$ -model could be having an undesirable effect on the final fit that, when available data are weaker, might be influencing the spatial patterns determined.

Regarding the fixed effects  $\mathbf{M}$ -model, we have mentioned that it was equivalent to assuming a  $Wishart(K, \sigma^2 \mathbf{I}_J)$  prior distribution on the covariance matrix between diseases  $\Sigma_b$ . Since  $\sigma$  is usually set to a large value for the fixed effects approach, this entails that the prior mean of  $\Sigma_b$  is equal to  $K\sigma^2 \mathbf{I}_J$ , for a high value of  $\sigma$ . Therefore, this model assumes the prior covariances between diseases to be 0 and prior variances of the log-risks within spatial patterns to be

high. These prior assumptions could explain the results found in Castellón for the fixed effects model, where the prior information in  $\mathbf{M}$  would overwhelm the information provided by the data. For this city, the cells of  $\Theta$  do not produce any smoothing in the risks fitted, as a consequence of their large prior variances (subsumed in matrix  $\mathbf{M}$ ), which does not produce any shrinkage. As a consequence, the smoothed SMRs estimated for this model reproduce the unsmoothed original SMRs that disease mapping models typically try to avoid.

The random effects  $\mathbf{M}$ -model also leads to a prior mean of  $K\sigma^2 I_J$  for  $\Sigma_b$  but with  $\sigma$  now being a parameter to be estimated within the model. This would potentially avoid the undesirable non-shrinking effect of the fixed effects  $\mathbf{M}$ -model when applied to smaller datasets. In this case the prior mean will just be proportional to the identity matrix but the proportionality constant will be estimated by the model itself, which will be set to a common consensus value for all the diseases. Univariate BYM models for each of the diseases in Castellón yielded posterior standard deviations for the log-SMRs ranging from 0.05 to 0.42, depending on the disease. AIDS was the disease with a higher standard deviation, far larger than the median standard deviation for the set of diseases considered (0.13). Thus, the distribution of the standard deviations of the log-SMRs for the different diseases has a pronounced asymmetrical right-tailed distribution. In consequence, the consensus scale parameter  $\sigma$  for the random effects model takes a value that is much lower than that required to appropriately describe the spatial variability of AIDS mortality. This could explain perfectly why the initial pattern highlighted by the univariate BYM model for AIDS vanishes when the random effects  $\mathbf{M}$ -model is fitted.

In sum, the Castellón multivariate mortality study above has shown important prior sensitivity for the  $\mathbf{M}$ -model, mainly for smaller datasets. Specifically, the fixed effects  $\mathbf{M}$ -model has a tendency to yield unsmoothed risk estimates. Furthermore, the random effects version has an inclination toward the shrinkage of all diseases to a common point in terms of variability. Although this could be fine for some particular datasets, in general it will be a restrictive performance of this model which it would be advisable to improve by seeking more adaptive models, at least in terms of the variance of the diseases. This is the goal that we pursue from now on.

## 4 An heteroscedastic modification of the $\mathbf{M}$ -model

Our proposal for fixing the prior sensitivity problems of the  $\mathbf{M}$ -model consists in a modification of its random effects version. Specifically, we relax the assumption of a common scale parameter for the cells of  $\mathbf{M}$ . In particular we propose two different ways to do this. The first proposal considers  $M_{ij} \sim N(0, \sigma_i^2)$  for  $i = 1, \dots, K$ , while our second alternative proposal considers  $M_{ij} \sim N(0, \sigma_j^2)$  for  $j = 1, \dots, J$ . From now on we will refer to these two proposals as the *row variance-adaptive* random effect  $\mathbf{M}$ -model (or simply RVA  $\mathbf{M}$ -model) and the *column variance-adaptive* random effect  $\mathbf{M}$ -model (or simply CVA  $\mathbf{M}$ -model),

respectively. Obviously these two proposals will be more adaptive in terms of variability than the original random effects  $\mathbf{M}$ -model, which will, hopefully, allow us to solve the shrinkage problems toward a common variability evidenced in the previous section. Henceforth, we will refer to the random effects  $\mathbf{M}$ -model introduced in Section 2 as simply the *non variance-adaptive* model (NVA model) in order to emphasize its main feature as compared to the two new variance-adaptive models that we have just introduced.

#### 4.1 An insight on the log-risks separation strategies for the RVA and CVA proposals

For all three RVA, CVA and NVA models  $\mathbf{M}$  can be stated as either  $\mathbf{D}\mathbf{M}^*$  or  $\mathbf{M}^*\mathbf{D}$  for  $\mathbf{D} = \text{diag}(\boldsymbol{\sigma})$  and  $M_{ij}^* \sim N(0, 1)$ , for  $\boldsymbol{\sigma}$  a vector of the appropriate length. Specifically,  $\mathbf{M} = \mathbf{D}\mathbf{M}^*$  for the RVA model,  $\mathbf{M} = \mathbf{M}^*\mathbf{D}$  for the CVA model, and  $\mathbf{M}$  can be stated as either  $\mathbf{M}^*\mathbf{D}$  or  $\mathbf{D}\mathbf{M}^*$  for  $\mathbf{D} = \sigma\mathbf{I}_J$  or simply  $\mathbf{M} = \sigma\mathbf{M}^*$  for the NVA model. This allows us to formulate the RVA model as

$$\boldsymbol{\Theta} = \boldsymbol{\Phi}\mathbf{D}\mathbf{M}^*, \quad (2)$$

or the CVA model as

$$\boldsymbol{\Theta} = \boldsymbol{\Phi}\mathbf{M}^*\mathbf{D} \quad (3)$$

in terms of Expression (1). In a similar manner,  $\boldsymbol{\Theta}$  in the original NVA model could also now be expressed as

$$\boldsymbol{\Theta} = \boldsymbol{\Phi}\mathbf{M}^*\boldsymbol{\sigma}, \quad (4)$$

or as both (2) and (3) for  $\mathbf{D} = \sigma\mathbf{I}_J$  instead of a general diagonal matrix as for RVA or CVA. We will use these expressions to further study the theoretical properties of these proposals instead of the RVA and CVA formulations in the first paragraph of this section. Although the formulation in that paragraph is more convenient in computational terms (indeed it has been the one used to implement these models in `WinBUGS`), the matrix formulations above are more convenient for studying the statistical properties of the corresponding models. So we will use them extensively from now on.

Expression (2) evidences an interesting interpretation of the RVA model, that is, the matrix decomposition there can also be viewed as  $(\boldsymbol{\Phi}\mathbf{D})\mathbf{M}^*$  and thus the standard deviations  $\boldsymbol{\sigma}$  in that model may be interpreted as those corresponding to the underlying spatial patterns in  $\boldsymbol{\Phi}$ . Hence, this model can be viewed as a set of underlying spatial patterns of different variability (in contrast to NVA) that are later made dependent by their postmultiplication by  $\mathbf{M}^*$ . On the other hand, the CVA model first makes the spatial patterns in  $\boldsymbol{\Phi}$  dependent (which originally had the same variability) and later those unscaled dependent patterns are scaled by means of the postmultiplication by  $\mathbf{D}$ . Therefore, the standard deviations  $\boldsymbol{\sigma}$  in both the RVA and CVA models have very different interpretations. First, for the RVA model, these standard deviations correspond to the underlying spatial patterns, whereas for the CVA model they scale the

(spatial and multivariate dependent  $\Phi M^*$ ) patterns available according to the variability needed for each particular disease.

Expressions (2) to (4) separate the different sources involved in the multivariate covariance structure into different terms. Similar separation strategies are also advocated by Barnard et al. (2000) in multivariate (non-spatial) problems and by MACNAB-SEIO in multivariate disease mapping studies. Our proposal runs in that same direction, with some advantages that we will describe below.

By this separation of  $\Theta$  into several components,  $\Phi$  is in charge of modeling the spatial dependence of the data,  $M^*$  is in charge of modeling the multivariate dependence between diseases, and  $D$  models the scale of  $\Theta$ . In any case, note that some confounding will remain between  $M^*$  and  $D$  since, ideally,  $M^*$  would be in charge of modeling the correlation matrix between diseases, but it does not do exactly that. To model the correlation matrix between diseases  $C_b$ ,  $M^*$  should be defined so that  $C_b = (M^*)'M^*$ . This would entail  $J$  column restrictions on  $M^*$ , specifically  $\{M_{.j}^*M_{.j}^* = 1 : j = 1, \dots, J\}$ , which are generally detrimental for MCMC algorithms (in our experience neither WinBUGS nor Stan tolerate restrictions of this kind very well). In contrast, we propose modeling  $M_{.j}^* \sim N(0, 1)$ . With this choice we have that  $M_{.j}^*M_{.j}^* \sim \chi_J^2$ , which will give  $J$  as expected value. Thus, our definition of  $M^*$  does not allow us to set  $M_{.j}^*M_{.j}^*$  to some specific value and therefore it will not model any correlation matrix. Nevertheless, the feature  $E(M_{.j}^*M_{.j}^*) = J$ , which makes  $J^{-1}(M^*)'M^*$  on average a correlation matrix, sets the scale of  $M^*M^*$ . This allows the modeling of the scale of the multivariate patterns to be separated into the separate matrix  $D$ , since that scale cannot be controlled by  $M^*$ .

Hence, we now have two alternative separation strategies that could fix the non-adaptability, in terms of variability, of the NVA proposal in Botella-Rocamora et al. (2015). We are now going to explore their differences for modeling  $\Theta$  through  $\Sigma_b$ .

## 4.2 An insight on the RVA and CVA proposals in terms of the modeling of $\Sigma_b$

The main difference between the RVA and CVA proposals lies in their inherently different ways of modeling  $\Sigma_b$ . Thus, for RVA  $\Sigma_b = M'M = (DM^*)'(DM^*) = (M^*)'D^2M^*$ , whereas for CVA  $\Sigma_b = (M^*D)'(M^*D) = D(M^*)'M^*D$ . According to these decompositions of  $\Sigma_b$  and expressions (2) and (3), RVA and CVA have markedly different interpretations. We start by analyzing RVA. Note that  $M^*$  in Expression (2) could be  $QR$ -decomposed as  $M^* = QR$  for suitable orthogonal ( $Q$ ) and upper triangular ( $R$ ) matrices. Therefore, Expression (2) could be alternatively stated as  $\Theta = \Phi DQR$ . If  $R = I_j$ , then, for RVA, we would have  $\Sigma_b$  being equal to  $(DM^*)'(DM^*) = (QR)'D^2QR = Q'D^2Q$ , that is,  $Q$  and  $D^2$  would contain the eigenvectors and eigenvalues, respectively, of  $\Sigma_b$ . Hence, in this case, we could interpret the RVA model as a PCA decomposition of  $\Theta$ , where  $\Phi$  would be the (spatially correlated) individual scores corresponding to each geographical unit,  $D$  would weight the contribution of



each axis to the multivariate dependence structure in  $\Theta$ , and  $Q$  contains the orthogonal axis defining the PCA. In the most general case in which  $R$  was not necessarily equal to  $I_j$ , the axis in the PCA would not be just  $Q$ , but  $QR = M^*$  and therefore, in that case, RVA can be understood as a PCA of  $\Theta$  followed by a subsequent non-orthogonal rotation. In this general case the PCA interpretation would therefore remain but with the original cloud of points projected onto a non-orthogonal axis given by  $M^*$ . The columns of  $\Phi$  could be understood as the individual scores corresponding to each spatial unit when projected onto those non-orthogonal components. The RVA model assumes that each of those columns corresponding to a specific linear combination of diseases (hopefully with a particular sense) follows a spatially structured distribution so, in some sense, RVA performs a non-orthogonal spatial PCA of the matrix of log-risks  $\Theta$ .

The non-orthogonal (spatially-correlated) PCA analysis performed in RVA could make more sense than it might seem at a first glance, since geographical patterns of risk factors would be rarely uncorrelated. Think, for example, of the spatial pattern of alcohol and tobacco consumption throughout a region of study. It would be hard to assume that both factors are independent. In that case, if these two risk factors were the two main determinants of the diseases in our study, a simple orthogonal PCA would induce spatially correlated distributions for both a linear combination of these factors (a weighted mean) and the corresponding orthogonal combination for these two variables. Assuming spatial distributions for these two components could not be justified since the second of them is mainly a residual shape component of the PCA, possibly showing weak spatial dependence. In contrast, a non-orthogonal PCA analysis, such as the one performed in the RVA model, would determine the same linear subspace for fitting the (correlated) effects but without assuming an orthogonal performance between alcohol- and tobacco-related spatial distributions. Thus, these two axes could focus on the separate geographical description of alcohol and tobacco consumption, when the assumption of spatial dependence for these two patterns is sure to be far more sensible than for the components of the regular orthogonal PCA.

Regarding CVA, Expression (3) could be alternatively stated as  $\Theta D^{-1} = \Phi M^* = \Phi Q R$  and, thus, CVA performs a matrix decomposition of the scale-standardized matrix  $\Theta D^{-1}$ . Since  $D^{-1} \Sigma_b D^{-1} = D^{-1} D (M^*)' M^* D D^{-1} = (M^*)' M^* = R' Q' Q R = R R'$  then  $R$  will correspond to the Cholesky upper triangle of the correlation matrix between diseases. In consequence, the columns of  $\Phi Q$  will correspond, respectively, to the individual scores explaining the (standardized) first disease, the individual scores explaining the (standardized) second disease given the first, and so forth. CVA assumes those vectors of scores ( $\Phi Q$ ) to be orthogonal combinations of common underlying spatial patterns. Those orthogonal combinations mean that all the columns of  $\Phi Q$  share a common distribution (all of them are linear combinations of the same spatial patterns) and this therefore makes the modeling of  $\Theta$  order-free with regard to diseases, i.e., invariant to their ordering (Martinez-Beneito, 2013).

A second interesting interpretation of the CVA model also comes from the decomposition  $\Sigma_b = \mathbf{D}(\mathbf{M}^*)'\mathbf{M}^*\mathbf{D} = \mathbf{D}\mathbf{W}\mathbf{D}$ , where  $\mathbf{W}$  follows a standard Wishart distribution  $Wishart(K, \mathbf{I}_J)$ . This is a scaled Wishart distribution as defined in Gelman et al. (2014). The scaled Wishart distribution has a clear advantage over the regular Wishart distribution as it separates the modeling of the variance parameters from that of the unscaled covariance structure. This allows it, for example, to be weakly informative on the scale parameters but more informative on the correlation structure of  $\Sigma_b$ , since being too uninformative on that structure makes the marginal priors of their correlation parameters accumulate most of its mass at their extremes. In contrast, assuming an inverse  $Wishart(J + 1, \mathbf{I}_J)$  distribution on  $\Sigma_b$ , which would mean putting flat prior distributions on its correlation parameters, assumes informative priors on its standard deviations (see page 286 in Gelman and Hill (2007)). Thus, the common degrees-of-freedom parameter of the Wishart distribution seems to introduce modeling conflicts between the correlation and standard deviation parameters of  $\Sigma_b$ . Mainly for these reasons, some authors advise the use of scaled Wishart priors instead of regular Wishart priors for modeling covariance/precision matrices (Barnard et al., 2000; Gelman and Hill, 2007; Gelman et al., 2014).

From a more practical point of view, the scaled Wishart distribution also allows specific inference to be performed on the different standard deviations  $\sigma_1, \dots, \sigma_J$  in a direct way. For a Wishart distribution, making inference on different standard deviations for each disease would require to increase the hierarchy of the model by setting  $Wishart(K, \mathbf{D})$  and putting an additional layer in the model for  $\mathbf{D}$ , if the software available allows us to do so. The most popular inference tools for spatial modeling nowadays (**WinBUGS** and **INLA**) have only implemented the Wishart distribution to model precision matrices in multivariate settings. Therefore, the proposed modeling overrides this limitation by building the scaled Wishart distribution by itself.

The Wishart and scaled Wishart distributions are frequently used as priors for precision matrices, instead of for covariance matrices, as we have assumed in our proposal. In our opinion the main reason for this consensus in the literature could be that the Wishart distribution is the conjugate distribution for precision matrices of multivariate Normal variables, what yields substantial benefits in computational and analytical terms. The use of the  $Wishart(J + 1, \mathbf{I}_J)$  as a prior distribution for precision matrices is particularly popular as it yields uniform marginal prior distributions on the correlation parameters between diseases (Barnard et al., 2000). Similarly, a scaled  $Wishart(J + 1, \mathbf{I}_J)$  prior distribution on the covariance matrix would mean a uniform prior distribution on the partial correlation parameters. Instead, our proposal puts a scaled  $Wishart(K, \mathbf{I}_J)$  distribution on the covariance matrix, which means that, for the common case  $K = J$ , this proposal should not be far from a uniform marginal prior distribution on the partial correlations between diseases. Nevertheless, Figure 2 illustrates the performance of our scaled  $Wishart(K, \mathbf{I}_J)$  prior for  $\Sigma_b$ , for  $K = J + 1$ , in terms of the correlations between diseases. Each graph

in that figure corresponds to the marginal distribution (histogram for 50,000 draws) of the correlation parameter between the first two diseases for scaled  $Wishart(J + 1, \mathbf{I}_J)$  prior distributions on  $\Sigma_b$ , for  $J = 3, 6, 12$  respectively. Figure 2 shows how the prior distributions for these settings concentrates on 0 as we increase the number of diseases. This is in contrast to the scaled  $Wishart(J + 1, \mathbf{I}_J)$  prior distribution on the precision matrix, which yielded uniform prior distributions on the marginal correlations independently of  $J$ , that is, the number of diseases considered. Results for other reasonable values of  $K$ , such as  $K = 2J$  (the value corresponding to the multivariate implementation of the BYM), depict similar performances although slightly more peaked toward 0.

Although the preference for small correlations might seem an undesirable effect for the scaled  $Wishart(K, \mathbf{I}_J)$  prior distribution on  $\Sigma_b$ , it could be more desirable than expected. As we increase the number of diseases, the number of marginal correlations between diseases in a model increases at a quadratic rate. Thus, assuming a uniform prior distribution for the marginal correlations between diseases would also mean a quadratic increase in false 'significant' correlations as the number of diseases increases. In contrast, a prior concentrating its mass on 0 when  $J$  grows would avoid this effect. This seems an interesting feature of our proposal as, when we increase the number of diseases in a multivariate study, we would expect the proportion of closely related diseases to go down instead of increasing at a quadratic rate. Hence, the prior structure proposed would perform a kind of multiplicity control on the number of related diseases (Scott and Berger, 2010), thereby inducing a parsimonious fit of the multivariate structure between them.

## 5 A new analysis of the Castellón mortality data

In this section we return to the geographical analysis of mortality in the city of Castellón and implement the new RVA and CVA variance-adaptive proposals described in the previous section. In order to evaluate those proposals, we compare the new estimated risks with those obtained with the NVA  $M$ -model and the univariate BYM models. The models have been executed in WinBUGS following the specifications introduced in Subsection 3.1. The R code for this analysis can also be found in the annex material, document `RCode.pdf`.

Figure 3 shows the estimated risk maps with the new modeling proposals for AIDS, Cerebrovascular disease and Suicides in men in Castellón. As can be seen in the case of AIDS, the new modeling proposals provide risk maps with greater variability than that obtained with the NVA model and closely similar to those estimated with the univariate BYM model (Figure 1). In the case of Cerebrovascular disease and Suicide, the risk maps estimated with the new modeling proposals present a considerable lower variability than the risk maps for AIDS. This shows that both RVA and CVA have solved the problem presented by the original multivariate NVA model, which provided risk maps

Table 1: DICs for the adjusted models in all three cities in the study

Model	Alicante	Castellón	Valencia
BYM with independent diseases	12964	6173	34270
Fixed effects $\mathbf{M}$ -model	13212	6675	34416
Random effects $\mathbf{M}$ -model	12865	6178	34029
RVA $\mathbf{M}$ -model	<b>12798</b>	<b>6148</b>	<b>33918</b>
CVA $\mathbf{M}$ -model	12870	6159	34009

with a similar variability for all the diseases in the study. Nevertheless, the original patterns estimated by the univariate BYM models seem to be reinforced for both the RVA and the CVA models (mainly for the RVA model), almost certainly as a consequence of sharing information between diseases. RVA and CVA estimates for Valencia and Alicante can also be found as annex material in the file `Valencia.VarAdapt.pdf` and `Alicante.VarAdapt.pdf`, respectively. Results for these cities confirm the visual conclusions also drawn for Castellón, although maybe to a lesser extent, since data for these cities are stronger than for Castellón.

Besides the visual comparison of the estimated risk maps with the different modeling proposals, we have also compared the fit of these models in general terms by using the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002). The DICs for all models and cities in the study can be found in Table 1. As can be observed, the model that provides a better fit in terms of the DIC in all three cities studied is the RVA  $\mathbf{M}$ -model, followed by the CVA  $\mathbf{M}$ -model in two out of three cities in the study. This seems to confirm that, besides the evident visual differences found, the heteroscedatic nature of the RVA and CVA models yields an important enhancement of the fit of the underlying geographical risk patterns.

## 6 Discussion

As described in this paper, the multivariate modeling proposal in Botella-Rocamora et al. (2015) for multivariate spatial studies of diseases presents some limitations when data are weaker. Specifically, in such situations, the prior structure of the  $\mathbf{M}$ -model can significantly influence the estimated risk patterns for all the diseases considered. As shown, this fact is caused by the single common variance parameter in the  $\mathbf{M}$  matrix of this model, which controls the overall variability of all risk patterns fitted. As illustrated, the fixed effects  $\mathbf{M}$ -model has a tendency to yield barely smoothed risk estimates as a consequence of assuming a high prior variance for the log-risks for all diseases. On the other hand, the random effects  $\mathbf{M}$ -model is prone to take all diseases, in terms of variability, to a common point that will be estimated by the model. If the variability of the risk patterns considered was different, these prior assumptions may produce evident misfits in the risk patterns that are estimated. One of the

main contributions of this work has been to highlight these limitations, which are particularly worrisome when the original NVA proposal is applied to small regions of study.

In this work we have proposed two modifications of the previous multivariate model that incorporate several different parameters to model the variability of the risks for each disease and which allow us to solve the problems evidenced in the study. These new heteroscedastic proposals allow the spatial patterns for each disease to have greater or lesser variability when necessary. This has made it possible to obtain more flexible and accurate risk estimates.

Regarding the two modeling proposals introduced in this paper, RVA has shown a better performance in empirical terms according to DIC. Thus, in practical terms, these proposals seem to be more advisable despite the appealing interpretation of the CVA model as a scaled Wishart prior on  $\Sigma_b$ . Beyond these empirical results we also find the RVA model proposed more interesting for several reasons. First, the PCA interpretation of the RVA approach seems quite interesting. Further work should be carried out under this approach in order to extract the 'principal maps' underlying this model since, as currently implemented, these factors cannot be identified by the model (some order restriction should be imposed, for example, in the vector of standard deviations  $\sigma$  in order to identify those 'principal maps'). Nevertheless, those 'principal maps' with the municipal scores corresponding to the different principal axes is an interesting idea that is certainly worth exploring. Moreover, the spatial modeling of the 'principal maps' that the RVA model makes is also appealing. Assuming a spatial distribution for these components could be a sensible assumption since these common underlying components could perfectly reflect the spatial distribution of risk factors throughout the region of study. In contrast, CVA assumes spatial distributions for the residual variability of each disease conditioned to the previous diseases in the study. We find this assumption much harder to support in practical terms. Nevertheless, the CVA proposal is interesting by itself because of its interpretation as a scaled Wishart prior for  $\Sigma_b$ . As shown, the CVA model makes it possible to implement the scaled Wishart within some regular Bayesian packages, such as WinBUGS, and this could be of interest even beyond the disease mapping literature.

## Acknowledgements

The authors acknowledge the support of the research grant PI16/01004 of Instituto de Salud Carlos III and predoctoral contract UGP-15-156 of FISABIO.

## References

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica*, 10:1281–1311.

- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–21.
- Botella-Rocamora, P., Martínez-Beneito, M. A., and Banerjee, S. (2015). A unifying modeling framework for highly multivariate disease mapping. *Statistics in Medicine*, 34(9):1548–1559.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC, Boca Raton, 3 edition.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press.
- Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.
- Macnab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: a general framework for coregionalized multivariate CAR models. 35(21):3827–3850.
- MacNab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: Order-dependent and order-free cMCARs. 25(4):1118–1144.
- Martínez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.
- Martínez-Beneito, M. A., Botella-Rocamora, P., and Banerjee, S. (2017). Towards a multidimensional approach to bayesian disease mapping. *Bayesian Analysis*, 12:239–259.
- Scott, J. G. and Berger, J. O. (2010). Bayesian and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Annals of Statistics*, 38(5):2587–2619.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and 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.

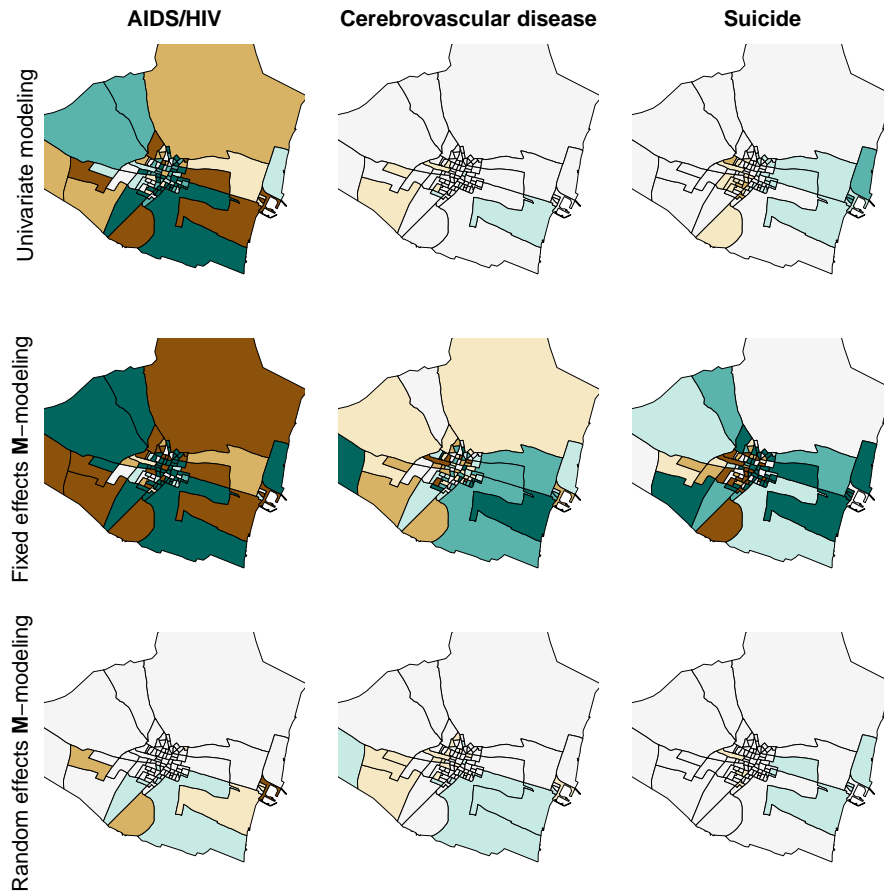


Figure 1: Choropleth maps for the estimated risk patterns using traditional univariate modeling (BYM), above, fixed effects  $M$ -modeling, center row, and random effects  $M$ -modeling, below

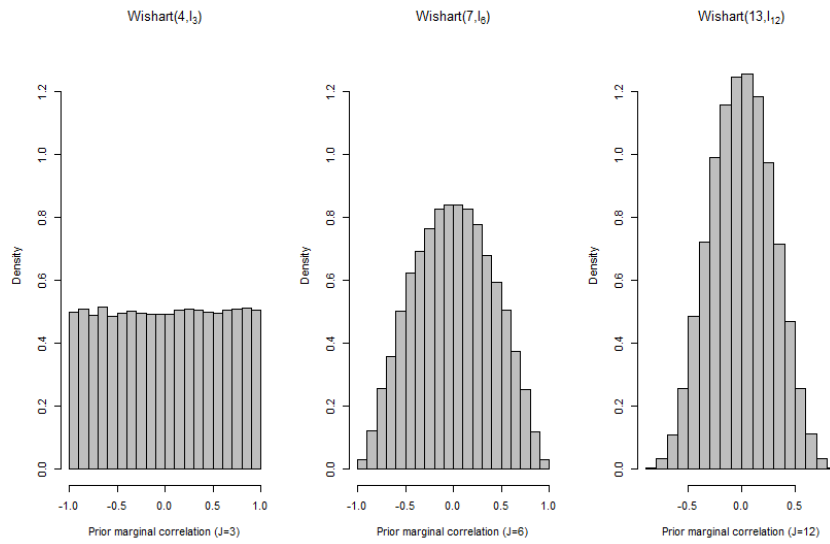


Figure 2: Prior marginal distributions for the correlation for the first two diseases, out of a set of  $J = 3, 6, 12$ , assuming a  $Wishart(J+1, \mathbf{I}_J)$  distribution for  $\Sigma_b$ . Histograms correspond to samples of 50,000 draws from the corresponding distribution of that marginal correlation



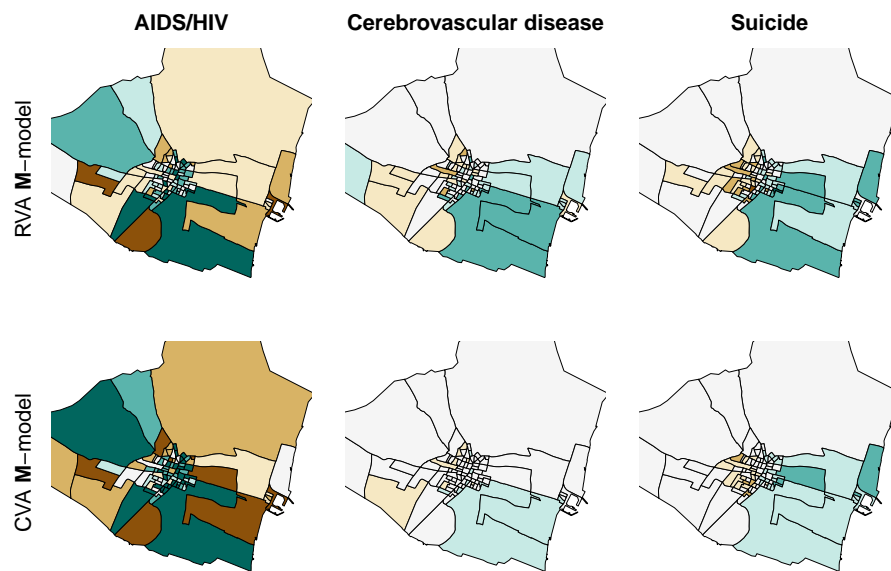


Figure 3: Choropleth maps for the estimated risks using the new heteroscedastic RVA and CVA  $M$ -models