Some links between conditional and coregionalized multivariate Gaussian Markov random fields*

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

Abstract

Multivariate disease mapping models are attracting considerable attention. Many modeling proposals have been made in this area, which could be grouped into three large sets: coregionalization, multivariate conditional and univariate conditional models. In this work we establish some links between these three groups of proposals. Specifically, we explore the equivalence between the two conditional approaches and show that an important class of coregionalization models can be seen as a large subclass of the conditional approaches. Additionally, we propose an extension to the current set of coregionalization models with some new unexplored proposals. This extension is able to reproduce asymmetric cross-spatial covariances for different diseases. This shows that the previously accepted belief that coregionalization was not able to reproduce models with asymmetric cross-covariances was wrong.

1 Introduction

Gaussian Markov random fields (GMRFs) are a family of multivariate distributions that make it possible to induce dependence structures on vectors of variables in a simple way. This has led to GMRFs being used in many different contexts, although they are particularly popular in disease mapping studies on sets of small areas. Recently, multivariate disease mapping has been the focus of considerable attention (Martinez-Beneito and Botella Rocamora, 2019). These models frequently use multivariate versions of Gaussian Markov random fields (MGMRFs) for the purpose of inducing both spatial and multivariate dependence (between diseases) for the joint study of several spatially referenced

---

*This is a preprint version of the paper: Martinez-Beneito, MA. Some links between conditional and coregionalized multivariate Gaussian Markov random fields. Spatial Statistics, 2020.

vectors, henceforth representing different diseases. In this manner, multivariate disease mapping models take advantage of the geographical patterns of hypothetically related diseases, which could be an important secondary source of information that regular (univariate) disease mapping studies typically ignore.

Multivariate disease mapping studies have been gaining interest in the literature over the last few years. As a result, many modeling proposals in this area have been put forward. Recently, MacNab (2018) proposed a classification of MGMRFs in the literature based on the way they induce spatial and multivariate dependence. This classification establishes three large groups of proposals: multivariate conditional models, mainly following the work of Mardia (1988); univariate conditional models, principally following the work of Sain et al. (2011), and coregionalization models (Gelfand et al., 2004; Jin et al., 2007; Martinez-Beneito, 2013). This classification includes most of the MGMRF-based modeling proposals already put forward in the literature. Besides the undeniable interest of MacNab's classification, it also raises some new questions (Martinez-Beneito, 2018), such as: Are the methods used to induce dependence in these three approaches complementary or are all three equivalent? Are any of these three approaches contained in some of the others or do all three show separate features that cannot be reproduced by the rest? Additionally, coregionalization models have sometimes been regarded as a minor family of multivariate models since they have been said to be incapable of reproducing asymmetric cross-spatial covariance matrices (covariance matrices for different diseases) (MacNab, 2018; Martinez-Beneito, 2018). This is in contrast to the mentioned conditional approaches, which have been able to reproduce those asymmetric settings (Sain and Cressie, 2007; Greco and Trivisano, 2009; Sain et al., 2011). These works also adduce several supposed advantages of those asymmetric models over their constrained symmetric alternatives.

The main goal of this work is to shed light on some of these questions. Particularly, we seek to establish some relationships among the classes of multivariate conditional, univariate conditional and coregionalization models. To do so, we define three sets of models based on this classification and we study the relationships among these three sets. Furthermore, when defining the set of coregionalized models, we formulate a new proposal within this approach that is more general than those already posited in the literature. This formulation is able to reproduce asymmetric cross-spatial dependencies, a feature which was not thought possible under this approach. In this manner, we increase the scope of the models that may be reproduced within this family. Finally, as an enhancement of the originally proposed asymmetric coregionalization models, we develop a particular proposal with much better computational properties, thereby showing its applicability in real large settings.

This paper is divided into 5 sections. Section 2 performs a brief review of the coregionalization approach and formulates the new general (asymmetric) model within this approach. Section 3 defines the sets of multivariate and univariate conditional models and explores the relationship between these two sets and coregionalization models. Section 4 applies the new coregionalization

model proposed in Section 2 to a real setting and, finally, Section 5 draws some conclusions from the results shown in this paper.

2 Coregionalized multivariate Gaussian Markov random fields

Coregionalization (Wackernagel, 2003; Gelfand et al., 2004) is one of the main approaches to MGMRFs. Coregionalization induces dependence among the cells of a random matrix, which will typically be used to model the log-risks for a set of diseases, as a product of several random matrices. Several proposals have been made under this approach for building multivariate GMRFs (Jin et al., 2007; Botella-Rocamora et al., 2015), which have been even used for multivariate joint studies of several sexes, time periods... (Martinez-Beneito et al., 2017). Some other proposals have also been formulated, merging both the conditional and the coregionalized approaches (MacNab, 2016, 2018), but henceforth we will focus on *pure coregionalization models* from independent underlying patterns, in order to establish a relationship between both approaches. To our knowledge, all (pure) coregionalized models formulated in the literature are particular cases of the QR-model by Martinez-Beneito (2013), which yields a covariance structure equivalent to that of Case 3 in Jin et al. (2007).

Let us now introduce the QR-model mentioned above. Let $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_{\cdot 1}, ..., \boldsymbol{\epsilon}_{\cdot J}]$ be an $I \times J$ random matrix with independent cells following a standard Normal distribution, where I denotes the number of spatial units available and J the number of diseases to be studied. Let us also consider $\boldsymbol{\phi} = [(\tilde{\boldsymbol{\Sigma}}_w)_1 \boldsymbol{\epsilon}_{\cdot 1}, ..., (\tilde{\boldsymbol{\Sigma}}_w)_J \boldsymbol{\epsilon}_{\cdot J}]$, for $(\tilde{\boldsymbol{\Sigma}}_w)_1, ..., (\tilde{\boldsymbol{\Sigma}}_w)_J$ being J Cholesky lower triangular matrices corresponding to the spatial covariance matrices $(\boldsymbol{\Sigma}_w)_1, ..., (\boldsymbol{\Sigma}_w)_J$ for each of the J diseases. Obviously, the columns of $\boldsymbol{\phi}$ will have different dependence structures given by the covariance matrices $(\boldsymbol{\Sigma}_w)_1, ..., (\boldsymbol{\Sigma}_w)_J$, which are supposed to induce different spatial patterns on each of these vectors. The QR-model proposes defining a multivariate GMRF $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_{\cdot 1}^T, ..., \boldsymbol{\zeta}_{\cdot J}^T)^T$ as the matrix product:

$$\boldsymbol{\zeta} = \boldsymbol{\phi}\hat{\boldsymbol{\Sigma}}_b^T = [(\tilde{\boldsymbol{\Sigma}}_w)_1 \boldsymbol{\epsilon}_{\cdot 1}, ..., (\tilde{\boldsymbol{\Sigma}}_w)_J \boldsymbol{\epsilon}_{\cdot J}]\hat{\boldsymbol{\Sigma}}_b^T,$$

where $\hat{\boldsymbol{\Sigma}}_b$ represents a matrix fulfilling $\boldsymbol{\Sigma}_b = \hat{\boldsymbol{\Sigma}}_b\hat{\boldsymbol{\Sigma}}_b^T$, for $\boldsymbol{\Sigma}_b$ a matrix reproducing the covariance between diseases. Martinez-Beneito (2013) shows that this definition does not uniquely identify $\hat{\boldsymbol{\Sigma}}_b$; in fact, the set of matrices fulfilling that relationship coincides with the set of orthogonal rotations of the Cholesky lower triangular matrix $\tilde{\boldsymbol{\Sigma}}_b$: $\{\tilde{\boldsymbol{\Sigma}}_b\boldsymbol{P} : \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}_J\}$. The QR-model yields the following joint covariance matrix for $vec(\boldsymbol{\zeta})$:

$$(\hat{\boldsymbol{\Sigma}}_b \otimes \boldsymbol{I}_I)Bdiag((\boldsymbol{\Sigma}_w)_1, ..., (\boldsymbol{\Sigma}_w)_J)(\hat{\boldsymbol{\Sigma}}_b^T \otimes \boldsymbol{I}_I),$$

where $Bdiag()$ denotes a block–diagonal matrix of the corresponding arguments. Note that if $\boldsymbol{\Sigma}_w := (\boldsymbol{\Sigma}_w)_1 = ... = (\boldsymbol{\Sigma}_w)_J$, that covariance matrix would be equal to simply $\boldsymbol{\Sigma}_b \otimes \boldsymbol{\Sigma}_w$, providing a separable covariance structure for $vec(\boldsymbol{\zeta})$.

3

Alternatively, it is also enlightening to consider the columns of $\boldsymbol{\phi}$ to follow proper CAR (pCAR) distributions of different correlation parameters $\rho_1, ..., \rho_J$. In that case, the covariance matrices for the columns of $\boldsymbol{\phi}$ would be simply $\{(\boldsymbol{\Sigma}_w)_j = (\boldsymbol{D} - \rho_j \boldsymbol{W})^{-1}\}_{j=1}^J$, with $\boldsymbol{W}$ the typical $I \times I$ adjacency matrix used for defining GMRFs and $\boldsymbol{D} = diag(w_{1+}, ..., w_{I+})$ with $w_{i+} = \sum_{i'=1}^I W_{ii'}$. It follows that, for this case, this covariance matrix for $vec(\boldsymbol{\zeta})$ could be expressed as:

$$(\hat{\boldsymbol{\Sigma}}_b \otimes \boldsymbol{I}_I)(Bdiag(\boldsymbol{D} - \rho_1 \boldsymbol{W}, ..., \boldsymbol{D} - \rho_J \boldsymbol{W}))^{-1}(\hat{\boldsymbol{\Sigma}}_b^T \otimes \boldsymbol{I}_I) =$$
$$(((\hat{\boldsymbol{\Sigma}}_b^T)^{-1} \otimes \boldsymbol{I}_I)(\boldsymbol{I}_J \otimes \boldsymbol{D} - diag(\boldsymbol{\rho}) \otimes \boldsymbol{W})(\hat{\boldsymbol{\Sigma}}_b^{-1} \otimes \boldsymbol{I}_I))^{-1} =$$
$$\left(\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{D} - [\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P} diag(\rho_1^{-1}, ..., \rho_J^{-1}) \boldsymbol{P}^T \tilde{\boldsymbol{\Sigma}}_b^T]^{-1} \otimes \boldsymbol{W} \right)^{-1},$$

where $\boldsymbol{P}$ stands for the orthogonal rotation that fulfills $\hat{\boldsymbol{\Sigma}}_b = \tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}$. By considering a general $\hat{\boldsymbol{\Sigma}}_b$ matrix in the QR-model formulation, instead of just the lower Cholesky matrix $\tilde{\boldsymbol{\Sigma}}_b$, the QR-model is no longer order-dependent and, therefore, the order considered for the diseases in the analysis becomes irrelevant (Martinez-Beneito, 2013). This is a particular advantage of QR-models.

2.1 An extension of the coregionalized QR-model

We now consider a variation of the QR-model as it has been defined in the previous paragraphs. First, as above, we induce different spatial patterns on the columns of the random matrix $\boldsymbol{\epsilon}$ by making $\boldsymbol{\phi} = [(\tilde{\boldsymbol{\Sigma}}_w)_1 \boldsymbol{\epsilon}_{\cdot 1}, ..., (\tilde{\boldsymbol{\Sigma}}_w)_J \boldsymbol{\epsilon}_{\cdot J}]$. However, dependence between diseases is induced in a different manner. Thus, if in the QR-model each row of $\boldsymbol{\phi}$ was post-multiplied by the same matrix $\hat{\boldsymbol{\Sigma}}_b^T$, we will now multiply each of its rows $\boldsymbol{\phi}_{i\cdot}$ by a different $(\hat{\boldsymbol{\Sigma}}_b)_i^T = (\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_i)^T = \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T$, which induces a common covariance structure between diseases $\boldsymbol{\Sigma}_b$, but depends on location-specific $J \times J$ permutation matrices $\boldsymbol{P}_1, ..., \boldsymbol{P}_I$. We will refer to this model henceforth as the QsR-model in order to stress the presence of several orthogonal matrices instead of just a single one, as was the case for the QR-model.

Let us now explore the covariance matrix of this ensemble for the particular case where we had pCAR distributions, of different parameters $\boldsymbol{\rho}$, for the set of underlying patterns $\boldsymbol{\phi}$. In that case we would have:

$$\boldsymbol{\Sigma}_{vec(\boldsymbol{\phi})} = Bdiag((\boldsymbol{D} - \rho_1 \boldsymbol{W})^{-1}, ..., (\boldsymbol{D} - \rho_J \boldsymbol{W})^{-1}) = (\boldsymbol{I}_J \otimes \boldsymbol{D} - diag(\boldsymbol{\rho}) \otimes \boldsymbol{W})^{-1}$$

or, equivalently (see Section 3 in Henderson and Searle (1979) or Ippoliti et al. (2018)):

$$\boldsymbol{\Sigma}_{vec(\boldsymbol{\phi}^T)} = (\boldsymbol{D} \otimes \boldsymbol{I}_J - \boldsymbol{W} \otimes diag(\boldsymbol{\rho}))^{-1}.$$

The second part of the former proposal, where multivariate dependence is induced, could be algebraically expressed as:

$$vec(\boldsymbol{\zeta}^T) = (vec(\boldsymbol{\phi}^T)^T Bdiag((\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_1)^T, ..., (\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_I)^T))^T$$
$$= (vec(\boldsymbol{\phi}^T)^T Bdiag(\boldsymbol{P}_1^T, ..., \boldsymbol{P}_I^T)(\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b^T))^T$$
$$= (\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b) Bdiag(\boldsymbol{P}_1, ..., \boldsymbol{P}_I) vec(\boldsymbol{\phi}^T).$$

Therefore, if $Block()$ denotes a block matrix of the corresponding arguments, then:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta}^T)} &= ((\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b) Bdiag(\boldsymbol{P}_1, ..., \boldsymbol{P}_I))(\boldsymbol{D} \otimes \boldsymbol{I}_J - \boldsymbol{W} \otimes diag(\boldsymbol{\rho}))^{-1} \\
&\quad \cdot (Bdiag(\boldsymbol{P}_1^T, ..., \boldsymbol{P}_I^T)(\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b^T)) \\
&= ((\boldsymbol{I}_I \otimes (\tilde{\boldsymbol{\Sigma}}_b^T)^{-1}) Bdiag(\boldsymbol{P}_1, ..., \boldsymbol{P}_I)(\boldsymbol{D} \otimes \boldsymbol{I}_J - \boldsymbol{W} \otimes diag(\boldsymbol{\rho})) \\
&\quad \cdot Bdiag(\boldsymbol{P}_1^T, ..., \boldsymbol{P}_I^T)(\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b^{-1}))^{-1} \\
&= ((\boldsymbol{I}_I \otimes (\tilde{\boldsymbol{\Sigma}}_b^T)^{-1})(\boldsymbol{D} \otimes \boldsymbol{I}_J - Block(\boldsymbol{P}_i diag(\boldsymbol{\rho}) \boldsymbol{P}_{i'}^T w_{ii'}))(\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b^{-1}))^{-1} \\
&= (\boldsymbol{D} \otimes \boldsymbol{\Sigma}_b^{-1} - Block((\tilde{\boldsymbol{\Sigma}}_b^T)^{-1} \boldsymbol{P}_i diag(\boldsymbol{\rho}) \boldsymbol{P}_{i'}^T (\tilde{\boldsymbol{\Sigma}}_b)^{-1} w_{ii'}))^{-1} \\
&= (\boldsymbol{D} \otimes \boldsymbol{\Sigma}_b^{-1} - Block((\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_{i'} diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1} w_{ii'}))^{-1}.
\end{aligned}
$$

If we denoted $\boldsymbol{Q}_{ii'} = (\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_{i'} diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1}$ and $\boldsymbol{e}_i$ the I-vector being 1 in its i-th component and 0 otherwise, it follows that $\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta}^T)}$ could be alternatively expressed as:

$$
\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta}^T)} = (\boldsymbol{D} \otimes \boldsymbol{\Sigma}_b^{-1} - Block(\boldsymbol{Q}_{ii'} w_{ii'}))^{-1} = (\boldsymbol{D} \otimes \boldsymbol{\Sigma}_b^{-1} - \sum_{i,i'=1}^{I} w_{ii'} \boldsymbol{e}_i \boldsymbol{e}_{i'}^T \otimes \boldsymbol{Q}_{ii'})^{-1}
$$

$$
= (\boldsymbol{D} \otimes \boldsymbol{\Sigma}_b^{-1} - \sum_{i,i'=1}^{I} \boldsymbol{W}_{ii'} \otimes \boldsymbol{Q}_{ii'})^{-1} \tag{1}
$$

where $\boldsymbol{W}_{ii'}$ denotes an $I \times I$ matrix with its (i, i')-th cell being equal to $w_{ii'}$ and the rest of the cells being 0. This expression also allows us to obtain the covariance matrix of $vec(\boldsymbol{\zeta})$ as:

$$
\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta})} = (\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{D} - \sum_{i,i'=1}^{I} \boldsymbol{Q}_{ii'} \otimes \boldsymbol{W}_{ii'})^{-1}. \tag{2}
$$

Note that this covariance matrix, for $\boldsymbol{P}_1 = ... = \boldsymbol{P}_I = \boldsymbol{P}$ is equal to simply:

$$
\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta})} = (\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{D} - \boldsymbol{Q} \otimes \boldsymbol{W})^{-1}
$$

for $\boldsymbol{Q} = (\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P} diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1}$, which coincides with $\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta})}$ for the QR-model in Martinez-Beneito (2013), which is therefore just a particular case of the current proposal.

We are going to look more closely at the covariance structure of the QsR-model, in particular the asymmetry of its non-diagonal blocks. The precision matrix for $vec(\boldsymbol{\zeta})$, arising as the inverse of Expression (2), has a particularly sparse structure. Specifically, it can be expressed as an $(IJ) \times (IJ)$ block matrix with $I \times I$ blocks. The (j, j')th block of this matrix, for $1 \le j, j' \le J$, is equal to $\boldsymbol{B}_{jj'} = (\boldsymbol{\Sigma}_b^{-1})_{jj'} \boldsymbol{D} - \sum_{i,i'=1}^{J} (\boldsymbol{Q}_{ii'})_{jj'} \boldsymbol{W}_{ii'}$, where $(\boldsymbol{Q}_{ii'})_{jj'}$ denotes the (j, j')-th

cell of $\boldsymbol{Q}_{ii'}$. This expression thus has an interesting consequence for $j \neq j'$ and $i \neq i'$:

$$(\boldsymbol{B}_{jj'})_{ii'} = -(\boldsymbol{Q}_{ii'})_{jj'} w_{ii'} = -((\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_{i'} diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1})_{jj'} w_{ii'}$$
$$\neq -((\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_i diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}_{i'}^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1})_{jj'} w_{i'i} = (\boldsymbol{B}_{jj'})_{i'i}.$$

Therefore, the cross-spatial covariance matrices of the QsR-model are asymmetric, in contrast to the QR-model and the rest of the pure multivariate coregionalized models previously proposed. This makes it possible to reproduce models with asymmetric cross-spatial covariance matrices within the coregionalized framework.

We find it relevant to comment that few modeling proposals with asymmetric cross-covariances have been proposed previously in the literature (Sain and Cressie, 2007; Greco and Trivisano, 2009; Sain et al., 2011). Nevertheless, a particular feature is present in most of them (MacNab, 2018). Specifically, for those previous asymmetric models the order used for the spatial units of analysis has been found to be relevant, when the order chosen is typically arbitrary. Specifically, Greco and Trivisano (2009) propose a model which, for any two i and i' with $1 \leq i < i' \leq I$ and $i \sim i'$, have $Cov(\boldsymbol{\zeta}_{i\cdot}, \boldsymbol{\zeta}_{i'\cdot}) = \tilde{\boldsymbol{B}}$ for an asymmetric matrix $\tilde{\boldsymbol{B}}$. The asymmetry of $\tilde{\boldsymbol{B}}$ makes $Cov(\zeta_{ij}, \zeta_{i'j'}) = \tilde{\boldsymbol{B}}_{jj'} \neq \tilde{\boldsymbol{B}}_{j'j} = Cov(\zeta_{ij'}, \zeta_{i'j})$ for any two $j \neq j'$. As a consequence, for any two neighboring spatial units i and i', the covariance between their j-th and j'-th component for $1 \leq j, j' \leq J$ is equal to either $\tilde{\boldsymbol{B}}_{jj'}$ or $\tilde{\boldsymbol{B}}_{j'j}$, depending only on whether $i' > i$ or $i' < i$. This label dependence, making the final fit sensitive to the ordering used for the spatial units, does not happen for the QsR-model. For this model, $Cov(\zeta_{ij}, \zeta_{i'j'})$ is different for any two spatial units i and i', as it depends specifically on $\boldsymbol{P}_i$ and $\boldsymbol{P}_{i'}$, so its cross-spatial covariances will not be restricted to taking just two values. Moreover, the ordering used for the diseases is irrelevant for the QsR-model, since by changing $\boldsymbol{P}_i$ by $\boldsymbol{P}_{i'}$, and vice versa, we could transpose the cross-covariance matrix for these two spatial units. In our opinion, this is an interesting property of QsR-models that makes them quite appealing.

Finally, we believe it worthwhile mentioning that the MGMRFs arising from QsR-models are valid by definition, since they are built as matrix operations involving full-rank matrices, which preserve the symmetry and positive definiteness of $\boldsymbol{\Sigma}_{vec(\boldsymbol{\phi}^T)}$. Consequently, the validity of $\boldsymbol{\Sigma}_{vec(\boldsymbol{\zeta}^T)}$ depends exclusively on the validity of $\boldsymbol{\Sigma}_{vec(\boldsymbol{\phi}^T)}$. This is not a minor advantage, since building models that fulfill these conditions is one of the main challenges in MGMRF modeling.

2.2 A computationally convenient QsR-model

The QsR-model, as formulated above, has an undeniable theoretical interest as it is able to reproduce order-free asymmetric coregionalized models. Nevertheless, that proposal introduces important drawbacks in practical terms. Specifically, the original QR-model has been reformulated in the literature by considering $\hat{\tilde{\boldsymbol{\Sigma}}}_b$ to be equal to a general $J \times J$ matrix $\boldsymbol{M}$ with independent normal cells (Botella-Rocamora et al., 2015). This model was called the M-model in contrast to the

original QR-model, which was parameterized as a function of an orthogonal matrix $\boldsymbol{P}$ and the Cholesky matrix of the covariance matrix between diseases $\tilde{\boldsymbol{\Sigma}}_b$. The great advantage of M-models is that they no longer need the direct use of orthogonal and Cholesky matrices and are therefore computationally much more convenient, making it possible to perform multivariate analyses on tens of diseases. Botella-Rocamora et al. (2015) showed the equivalence of the QR and M-models, although, as mentioned, they are very different in practical terms.

The QsR-proposal posed above as an M-model would require the use of a different $\boldsymbol{M}$ matrix per spatial unit, but all of them would have to be of the form $\boldsymbol{M}_i = (\hat{\boldsymbol{\Sigma}}_b)_i^T = \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T$ for different orthogonal matrices $\boldsymbol{P}_i$. Sampling from this set of matrices would require working back again with orthogonal and Cholesky matrices and so the computational advantage of the M-model would be lost for the general QsR-model. However, the computational benefit of the M-model is something that we would not want to lose with the above proposal of coregionalized asymmetric models. Thus, our current goal is to formulate a variation of the QsR-model with the computational advantages of the M-model.

Asymmetric QsR-models arise as the outcome of using a different orthogonal matrix $\boldsymbol{P}_i$ for each spatial unit but a common covariance structure between diseases $\boldsymbol{\Sigma}_b$. Obviously, this general asymmetric model is over-parameterized, since a different orthogonal matrix should be estimated per spatial unit. Thus, we could formulate a more convenient modeling proposal by restricting the orthogonal matrices $\boldsymbol{P}_i$ to a particular subset of the whole set of $J \times J$ orthogonal matrices. As mentioned previously, the general QsR-model can be formulated as:

$$
\begin{aligned}
vec(\boldsymbol{\zeta}^T) &= (\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b) Bdiag(\boldsymbol{P}_1, ..., \boldsymbol{P}_I) vec(\boldsymbol{\phi}^T) \\
&= Bdiag(\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_1, ..., \tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_I) vec(\boldsymbol{\phi}^T) = Bdiag(\boldsymbol{M}_1^T, ..., \boldsymbol{M}_I^T) vec(\boldsymbol{\phi}^T),
\end{aligned}
$$

while the QR-model had a single $\boldsymbol{P}$ matrix which also yielded a common $\boldsymbol{M}$ matrix. We could simplify the general QsR-model as follows. Let us consider the orthogonal matrices above to be of the form $\boldsymbol{P}_i = \boldsymbol{P} \boldsymbol{R}_i$ for $\boldsymbol{P}$, a common orthogonal matrix for all spatial units, and $\boldsymbol{R}_i$ a collection of $J \times J$ permutation (and therefore orthogonal) matrices. The collection of matrices $\{\boldsymbol{P}_i; i = 1, ..., I\}$, as just defined, are a subset of the set of $J \times J$ orthogonal matrices, and so this model is a particular (restrictive) case of the general QsR-model, henceforth a *permuted QsR-model*.

Obviously, the permuted QsR-model is less parameterized than the general QsR-model. Algebraically, the permuted QsR-model can be expressed as:

$$
\begin{aligned}
vec(\boldsymbol{\zeta}^T) &= (\boldsymbol{I}_I \otimes \tilde{\boldsymbol{\Sigma}}_b) Bdiag(\boldsymbol{P}\boldsymbol{R}_1, ..., \boldsymbol{P}\boldsymbol{R}_I) vec(\boldsymbol{\phi}^T) \\
&= Bdiag(\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}\boldsymbol{R}_1, ..., \tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}\boldsymbol{R}_I) vec(\boldsymbol{\phi}^T) \\
&= Bdiag(\boldsymbol{M}^T \boldsymbol{R}_1, ..., \boldsymbol{M}^T \boldsymbol{R}_I) vec(\boldsymbol{\phi}^T) \\
&= Bdiag((\boldsymbol{R}_1^T \boldsymbol{M})^T, ..., (\boldsymbol{R}_I^T \boldsymbol{M})^T) vec(\boldsymbol{\phi}^T).
\end{aligned}
$$

This means that, for the permuted QsR-model, a different matrix $\boldsymbol{M}_i = \boldsymbol{R}_i^T \boldsymbol{M}$ is used to induce multivariate dependence in $\boldsymbol{\zeta}$ and these matrices are simple

copies of a common $\boldsymbol{M}$ matrix with their rows permuted in a different manner for each spatial unit, according to the corresponding permutation matrix $\boldsymbol{R}_i^T$. Therefore, sampling from this proposal avoids the need to sample from the space of orthogonal matrices $\boldsymbol{P}_i$ and the space of Cholesky matrices $\tilde{\boldsymbol{\Sigma}}_b$, and so the computational advantages of the M-model still remain in the permuted QsR-model. In any case, the permuted QsR-model is able to reproduce asymmetric cross-covariance matrices like the original QsR-model.

Note that although the permuted QsR-model is just a particular restrictive case of the general QsR-model, the set of permuted QsR-models is much larger than the original QR-model. Specifically, for J diseases there are $J!$ possible choices for each $\boldsymbol{R}_i$ and, therefore, for I spatial units and given $\boldsymbol{\phi}$ and $\boldsymbol{M}$, the permuted QsR-model is able to reproduce $(J!)^I$ different combinations of the underlying patterns. For just $J = 2$ diseases and a moderate region of study, this number could easily be quite large. Thus, the permuted QsR-model introduces considerable additional flexibility in the modeling. For even a moderate number of diseases, the permuted QsR-model would contain a huge number of variations as compared to the original QR-model, perhaps also causing it to be over-parameterized. In that case, a subset of all possible $J!$ permutations of the rows of $\mathbf{M}$ could be more convenient than considering all the possible permutations. This would reduce the complexity of the permuted QsR-model, thereby possibly yielding a more parsimonious fit. We will also consider this possibility in Section 5.

3   Conditional approaches to multivariate Gaussian Markov random fields

Defining MGMRFs through the definition of the conditional distributions of the cells of $\boldsymbol{\zeta}$ is the main alternative to coregionalization models. In this section we introduce two classes of multivariate and univariate conditional MGMRFs, respectively. Later on, we will examine the links between these two classes of models and then explore some relationships between them and two additional classes of coregionalized models to be defined below.

3.1   The class of multivariate conditional MGMRFs ($\mathcal{C}_M$):

Mardia (1988) proposes a way of generalizing univariate GMRFs to multivariate studies. His work starts with the original formulation of univariate GMRFs $\boldsymbol{\theta}$ by Besag (1974) as a set of univariate conditional distributions on each of its components $\theta_i | \boldsymbol{\theta}_{-i}$. Specifically, for the zero-mean case, the conditional formulation of univariate GMRFs correspond to the model arising from the set of conditional distributions:

$$\theta_i | \boldsymbol{\theta}_{-i} \sim N \left(\sum_{i' \sim i} \beta_{ii'} \theta_{i'}, \sigma_i^{-2} \right), \ i = 1, ..., I,$$

where $i' \sim i$ denotes all subindexes i' corresponding to neighboring spatial units of i, and $\boldsymbol{\theta}_{-i}$ stands for the vector $\boldsymbol{\theta}$ leaving out its i-th component. Note that we parameterize this normal distribution and the rest in this paper in terms of their precision or precision matrices. Some restrictions are imposed on $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ so that this set of conditional distributions yields a valid joint multivariate distribution for $\boldsymbol{\theta}$.

In a similar manner, Mardia's proposal contains all those models which can be expressed in the following form. For an $(I \times J)$-matrix $\boldsymbol{\zeta}$, let us consider the multivariate conditional distributions of each of its rows given the rest as:

$$\boldsymbol{\zeta}_{i.}|vec(\boldsymbol{\zeta}_{-i.}^T) \sim N_J \left(\sum_{i' \sim i} \boldsymbol{\beta}_{ii'} \boldsymbol{\zeta}_{i'.}^T, \boldsymbol{\Omega}_i \right) \quad i = 1, ..., I \tag{3}$$

where $\{\boldsymbol{\beta}_{ii'} : i, i' = 1, ..., I\}$ stands for a set of general unrestricted $J \times J$ matrices, $\{\boldsymbol{\Omega}_i : i = 1, ..., I\}$ is a collection of $J \times J$ symmetric, positive definite precision matrices. Note that, although Mardia originally defines the mean of the conditional distributions above as a sum for $i' \neq i$, it is usually assumed that $\boldsymbol{\beta}_{ii'} = \boldsymbol{0}$ for $i \not\sim i'$, as we have done, in order to induce Markovianity in the overall process. Furthermore, for notational convenience, $\boldsymbol{\beta}_{ii}, i = 1, ..., I$ are assumed to be equal to $-\boldsymbol{I}_J$ (Mardia, 1988). Obviously, as for univariate GMRFs, some requirements will have to be imposed on $\boldsymbol{\beta}_{ii'}$ and $\boldsymbol{\Omega}_i$ in order to guarantee the symmetry and positive definiteness of the overall precision matrix $\boldsymbol{\Omega}_{vec(\boldsymbol{\zeta})}$. Nevertheless, we do not develop those conditions further as they will not be of particular interest throughout the rest of this paper. We will refer to this class of models, for any particular choice of $\boldsymbol{\beta}_{ii'}$ and $\boldsymbol{\Omega}_i$, as $\mathcal{C}_M$, where the subindex M stands for *multivariate* conditional models, i.e. MGMRFs defined through the multivariate conditional distribution of the observations for each location given the rest. See, for example, Billheimer et al. (1997); Carlin and Banerjee (2003); Gelfand and Vounatsou (2003); Sain and Cressie (2007) and Greco and Trivisano (2009) for some particular proposals in the literature belonging to $\mathcal{C}_M$.

3.2 The class of univariate conditional MGMRFs ($\mathcal{C}_U$):

We are now going to define a second class of conditional models that do not belong to $\mathcal{C}_M$. This class of models corresponds to the set of univariate conditional models that would be particular cases of the general modeling proposal in Sain et al. (2011). Specifically, this set of MGMRFs contains all those models which can be expressed in the following form. For an $(I \times J)$-matrix $\boldsymbol{\zeta}$, let us consider the univariate conditional distributions of each of its cells given the rest as:

$$\zeta_{ij}|vec(\boldsymbol{\zeta}_{-(ij)}) \sim N \left(\sum_{i' \sim i} \beta_{iji'j} \zeta_{i'j} + \sum_{j' \neq j} \beta_{ijij'} \zeta_{ij'} + \sum_{i' \sim i} \sum_{j' \neq j} \beta_{iji'j'} \zeta_{i'j'}, \sigma_{ij}^2 \right)$$

for a general $I \times J \times I \times J$ 4-dimensional array $\boldsymbol{\beta}$. Some requirements will also have to be imposed for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ in order to guarantee the validity of $\boldsymbol{\Omega}_{vec(\boldsymbol{\varsigma})}$,

9

although, once again, they will not be of particular interest henceforth. We will refer to this set of models as $\mathcal{C}_U$, where the subindex U stands for *univariate* conditional models, i.e. MGMRFs defined through the multivariate conditional distribution of each of its cells given the rest. See, for example, Kim et al. (2001); Jin et al. (2007) and Sain et al. (2011) for some particular cases of models in $\mathcal{C}_U$.

Thus, to sum up, we have two classes of models $\mathcal{C}_M$ and $\mathcal{C}_U$ available, defined as collections of conditional models. The first set relies on the multivariate conditional distributions of the rows of a matrix $\boldsymbol{\zeta}$ given the rest of the rows, while the second relies on the univariate conditional distributions of each cell of $\boldsymbol{\zeta}$ given the rest of the cells.

3.3   Multivariate versus univariate conditional models

We are now going to explore the relationship between the two sets of conditional MGMRFs defined above. Ippoliti et al. (2018) have recently pointed out the equivalence of these two sets, that is, that any model in $\mathcal{C}_M$ can be equivalently reexpressed as a model in $\mathcal{C}_U$, and vice versa. The following two results (see the final Annex for their proofs) shed some more light on that equivalence. In particular they provide the particular equivalence of any of the MGMRFs in any of these classes as an element of the other class.

Theorem 1 *Let $\boldsymbol{\zeta}$ be an MGMRF in $\mathcal{C}_M$ defined as:*

$$\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}\boldsymbol{\zeta}_{i'\cdot}^T, \boldsymbol{\Sigma}_i^{-1}), i = 1, ..., I,$$

then $\boldsymbol{\zeta}$ also belongs to $\mathcal{C}_U$. In particular:

$$\zeta_{ij}|vec(\boldsymbol{\zeta}_{-(ij)}) \sim N\left(\sum_{i'\sim i}\beta_{iji'j}^*\zeta_{i'j} + \sum_{j'\neq j}\beta_{ijij'}^*\zeta_{ij'} + \sum_{j'\neq j}\sum_{i'\sim i}\beta_{iji'j'}^*\zeta_{i'j'}, \sigma_{ij}^{-2}\right)$$

for $\beta_{iji'j'}^* = (diag(\boldsymbol{\Sigma}_i^{-1})^{-1}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_{ii'})_{jj'}$ and $\sigma_{ij}^{-2} = (\boldsymbol{\Sigma}_i^{-1})_{jj}$.

As a main consequence of this result, we have that any multivariate conditional model, as defined above, can also be expressed as a univariate conditional model and therefore $\mathcal{C}_M \subseteq \mathcal{C}_U$. Theorem 1 also provides a way of expressing any multivariate conditional model as an univariate conditional model. However, the following theorem gives additional insight into the relationship between these two sets.

Theorem 2 *Given $\boldsymbol{\zeta}$, an MGMRF in $\mathcal{C}_U$, $\boldsymbol{\zeta}$ can also be formulated as a multivariate conditional model in $\mathcal{C}_M$. Specifically, if:*

$$\zeta_{ij}|vec(\boldsymbol{\zeta}_{-(ij)}) \sim N\left(\sum_{i'\sim i}\beta_{iji'j}\zeta_{i'j} + \sum_{j'\neq j}\beta_{ijij'}\zeta_{ij'} + \sum_{j'\neq j}\sum_{i'\sim i}\beta_{iji'j'}\zeta_{i'j'}, \sigma_{ij}^{-2}\right)$$

then this model can also be equivalently expressed as:

$$\boldsymbol{\zeta}_{-i.}|vec(\boldsymbol{\zeta}_{-i.}^T) \sim N_J\left(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}^*\boldsymbol{\zeta}_{i'.}^T, diag(\boldsymbol{\sigma}_{i.}^{-2})\right), \; i=1,...,I. \qquad (4)$$

for $\boldsymbol{\beta}_{ii'}^* = (\beta_{iji'j'})_{j,j'=1}^J$.

As a consequence of this result $\mathcal{C}_U \subseteq \mathcal{C}_M$ and therefore, from Theorems 1 and 2, we have that $\mathcal{C}_M = \mathcal{C}_U$, which has also been recently stated by Ippoliti et al. (2018) as mentioned above. In other words, any multivariate conditional model can be alternatively formulated as an equivalent univariate model, and vice versa. So the preference for posing the multivariate or the univariate version of any conditional model should just be a question of convenience in terms of software availability, sparser implementations of any of these alternatives, better convergence of MCMC algorithms, and so forth. Moreover, Theorems 1 and 2 specify for any model in both $\mathcal{C}_M$ and $\mathcal{C}_U$ its alternative formulation as elements of the other set of models. This alternative formulation might be helpful, for example, to determine conditions on the validity of any conditionally formulated MGMRF since that alternative formulation could perhaps yield useful conditions based on the specific theory of that class of models.

Theorems 1 and 2 also have an interesting consequence for multivariate conditional models. Thus, any model in $\mathcal{C}_M$ can be alternatively expressed as another equivalent multivariate conditional model with $\boldsymbol{\Omega}_i$ diagonal for any i. This is a direct consequence of applying Theorems 1 and 2, consecutively, to any model in $\mathcal{C}_M$. Specifically, any multivariate conditional model formulated as Expression (3) can be equivalently formulated as Expression (4) with $\boldsymbol{\beta}_{ii'}^* = diag(\boldsymbol{\Omega}_i)^{-1}\boldsymbol{\Omega}_i\boldsymbol{\beta}_{ii'}$ and $\boldsymbol{\sigma}_{ij}^2 = (\boldsymbol{\Omega}_i)_{jj}$. Therefore, the subclass of elements of $\mathcal{C}_M$ with diagonal conditional precision matrices $\boldsymbol{\Omega}_i$ for all i is as rich as the whole $\mathcal{C}_M$, since it contains models with joint covariance structures that are identical to any model in $\mathcal{C}_M$. As a consequence, considering general variance-covariance matrices $\boldsymbol{\Omega}_i$ in $\mathcal{C}_M$ seems unnecessary and it would suffice to consider $\boldsymbol{\Omega}_i$ to be diagonal matrices in order to generate all the covariance structures considered in $\mathcal{C}_M$.

3.4 Coregionalized versus conditional models

We are now going to explore the relationship between the classes of pure coregionalized and conditional models, $\mathcal{C}_U$ and $\mathcal{C}_M$, as defined above. We start by first defining two classes of coregionalized multivariate models and we will later explore their relationship with the classes of conditional models.

3.4.1 The classes of coregionalized MGMRFs ($\mathcal{C}_C$ and $\mathcal{C}_C^*$):

We consider as the set of coregionalized models that composed of the *QsR*-model, as defined in the previous section, and all its particular cases. As mentioned, the covariance matrix of that general model is that of Expression (2),

which has as (symmetric) particular cases the QR-model and M-models already mentioned, the Case 3 model in Jin et al. (2007) or that in Gelfand and Vounatsou (2003), to cite just some particular cases. Obviously this set also contains the permuted QsR-model. We will refer to this set of models as $\mathcal{C}_C$, where the subindex C stands for *coregionalized* models.

Besides the set $\mathcal{C}_C$, there is a second extended set of coregionalization models that could be of interest for this discussion. The models in $\mathcal{C}_C$, as introduced in Section 2, rely on underlying pCAR distributions. Nevertheless, nothing prevents the underlying patterns ϕ of coregionalized models from having more general spatial structures beyond the GMRF framework (see Knorr-Held and Raßer (2000) or Botella-Rocamora et al. (2012) for two particular examples). In that case we could define an extended set of coregionalization models $\mathcal{C}_C^*$ as that arising when general spatial structures, beyond GMRF, are considered to build the asymmetric model introduced in Section 2. We do not discuss the elements of this set in more detail here as it is not required for the subsequent discussion of the relationship between classes of models. Note that, to our knowledge, $\mathcal{C}_C^*$ contains as particular cases all the pure coregionalization models, with independent columns in ϕ, already proposed in the literature.

3.4.2 Some links between coregionalized and conditional models

Theorem 2.2 in Mardia (1988) allows us to derive for any MGMRF the multivariate conditional distributions $\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T)$ from the joint distribution of $vec(\boldsymbol{\zeta}^T)$. For the zero-centered case, that theorem states that if $vec(\boldsymbol{\zeta}^T) \sim N_{IJ}(\mathbf{0}_{IJ}, \boldsymbol{\Sigma}^{-1})$ for a block-precision matrix $\boldsymbol{\Sigma}^{-1} = Block(\{(\boldsymbol{\Sigma}_{ii'}^{-1})_{i,i'=1}^I\})$ of $J \times J$ matrices, then $\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii}^{-1})$, with:

$$\boldsymbol{\mu}_i = -\sum_{i'\sim i}(\boldsymbol{\Sigma}_{ii}^{-1})^{-1}\boldsymbol{\Sigma}_{ii'}^{-1}\boldsymbol{\zeta}_{i'\cdot}^T.$$

It is easy to check that this result, when applied to the QsR-model with pCAR distributions for ϕ, which had $\boldsymbol{\Sigma}_{ii'}^{-1} = D_{ii'}\boldsymbol{\Sigma}_b^{-1} - w_{ii'}\boldsymbol{Q}_{ii'}$ (see Expression (1)), yields the following set of multivariate conditional distributions:

$$\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}\boldsymbol{\zeta}_{i'\cdot}^T, D_{ii}\boldsymbol{\Sigma}_b^{-1}) \tag{5}$$

where D_{ii}, the i-th diagonal element of $\boldsymbol{D}$, is just the number of neighbors of the i-th spatial unit if adjacency is considered as neighboring criterion for the pCAR distribution, and $\boldsymbol{\beta}_{ii'} = -(\boldsymbol{\Sigma}_{ii}^{-1})^{-1}\boldsymbol{\Sigma}_{ii'}^{-1}$, which for $i' \sim i$ is:

$$\boldsymbol{\beta}_{ii'} = -(D_{ii}^{-1}\boldsymbol{\Sigma}_b)(-w_{ii'}\boldsymbol{Q}_{ii'}) = (D_{ii}^{-1}w_{ii'})\boldsymbol{\Sigma}_b(\tilde{\boldsymbol{\Sigma}}_b\boldsymbol{P}_{i'}diag(\boldsymbol{\rho}^{-1})\boldsymbol{P}_i^T\tilde{\boldsymbol{\Sigma}}_b^T)^{-1}$$
$$= (D_{ii}^{-1}w_{ii'})\tilde{\boldsymbol{\Sigma}}_b\boldsymbol{P}_idiag(\boldsymbol{\rho})\boldsymbol{P}_{i'}^T\tilde{\boldsymbol{\Sigma}}_b^{-1}.$$

Slightly different expressions would be obtained for ϕ following another univariate GMRFs beyond pCAR. As a consequence, any model in $\mathcal{C}_M$ of the kind

above will also be an element of $\mathcal{C}_C$, that is, it could be alternatively formulated as a coregionalization model. Conversely, any conditional model that cannot be expressed as the expressions above for suitable $\boldsymbol{P}_1, ..., \boldsymbol{P}_I$ and $\boldsymbol{\Sigma}_b$ matrices will not belong to $\mathcal{C}_C$.

In a similar manner, Theorem 1 in the Annex applied to Expression (5) yields that the QsR-model can be expressed as an univariate conditional model in $\mathcal{C}_U$ with

$$\beta_{iji'j'} = (D_{ii}^{-1} w_{ii'} diag(\boldsymbol{\Sigma}_b^{-1})^{-1})^{-1} (\tilde{\boldsymbol{\Sigma}}_b \boldsymbol{P}_{i'} diag(\boldsymbol{\rho}^{-1}) \boldsymbol{P}_i^T \tilde{\boldsymbol{\Sigma}}_b^T)^{-1})_{jj'} \qquad (6)$$

and $\sigma_{ij}^{-2} = D_{ii} diag(\boldsymbol{\Sigma}_b^{-1})_j$. Thus, any model in $\mathcal{C}_U$ with $\beta_{iji'j'}$ and σ_{ij}^{-2} of this kind also belongs to $\mathcal{C}_C$ and can therefore take advantage of the computational and theoretical results of that approach. Conversely, if an univariate conditional model could not take the expression above for suitable $\boldsymbol{P}_1, ..., \boldsymbol{P}_I$ and $\boldsymbol{\Sigma}_b$ matrices, it will not be an element of $\mathcal{C}_C$.

The results above are quite interesting as they show that the most general case in $\mathcal{C}_C$, and therefore all models in that family, can be expressed as multivariate conditional models of $\mathcal{C}_M$ and $\mathcal{C}_U$, and as a consequence $\mathcal{C}_C \subseteq \mathcal{C}_M = \mathcal{C}_U$. However, $\mathcal{C}_M$ or $\mathcal{C}_U$ are not contained in $\mathcal{C}_C$ as some of their models cannot be reproduced as QsR-models. For example, let us consider the multivariate conditional model of Greco and Trivisano (2009), which belongs to $\mathcal{C}_M$. Their proposal, as mentioned previously, considered the same asymmetric cross-covariance matrix $\tilde{\boldsymbol{B}}$ for all neighboring units i and i', with $i' > i$. This will not be possible for QsR-models since if $\boldsymbol{Q}_{ii'} = \boldsymbol{Q}_{ii''}$ for all neighboring units i, i' and i'' for $1 \leq i < i', i'' \leq I$, then $\boldsymbol{P}_1 = ... = \boldsymbol{P}_I$, which would necessarily yield symmetric (and identical) cross-covariance matrices for any two pairs of spatial units. In consequence, the former relationship between $\mathcal{C}_C$ and $\mathcal{C}_M$ is really of the kind $\mathcal{C}_C \subset \mathcal{C}_M = \mathcal{C}_U$ instead of just $\mathcal{C}_C \subseteq \mathcal{C}_M = \mathcal{C}_U$.

Expressions (5) and (6) also have a second interesting interpretation. QsR-models, viewed as such Expressions, yield interesting subclasses of $\mathcal{C}_M$ and $\mathcal{C}_U$ with valid models by construction, when this condition is not so easy to achieve. Indeed, for these models, necessary and sufficient conditions on $\boldsymbol{\rho}$ exist so that coregionalized models produce valid proposals. The search for valid elements in $\mathcal{C}_M$ or $\mathcal{C}_U$ is one of the main challenges when proposing new multivariate conditional models, and sufficient (but not necessary) conditions on $\boldsymbol{\rho}$ are usually available for those models. Expressions (5) and (6) characterize the elements of $\mathcal{C}_M$ and $\mathcal{C}_U$ that are also elements of $\mathcal{C}_C$. Accordingly, necessary and sufficient conditions on $\boldsymbol{\rho}$ can be formulated for them to produce valid models. Thus, for these cases we will not miss any model as a consequence of setting only sufficient conditions that could lead us to consider just a restrictive subset of models instead of all the possible valid models.

Finally, we consider it appropriate to conclude by mentioning that the set of coregionalized models is really larger than might be expected according to the above results. We must recall that we have defined a second extended set of coregionalized models $\mathcal{C}_C^*$ which also contained coregionalized models relying on non-GMRF underlying structures. Therefore, coregionalized models are able to generate multivariate spatial dependence structures which will not necessarily be

13

MGMRFs. In contrast, multivariate conditional models, as sets of conditional normal distributions, define joint normal distributions which are necessarily MGMRFs. Consequently, any model in $\mathcal{C}_C^* \backslash \mathcal{C}_C$ will not belong to $\mathcal{C}_M = \mathcal{C}_U$, thereby confirming that the set of coregionalized models is larger than might initially be expected.

4   An application of the permuted *QsR*-model

The aim of proposing *QsR*-models in this work is mainly theoretical, to make the family of coregionalized models larger. Nevertheless, we think it is worthwhile illustrating the feasibility and performance of the permuted *QsR*-model (its computationally convenient version) in a real setting. Accordingly, we have performed a three-variate mortality study in the Valencian Region, one of the 17 regions that Spain is divided up into. In this study we have analyzed cirrhosis, lung cancer, and oral cancer mortality for males, during the period 1991-2011, at the municipal level (a total of 540 municipalities). For the study of these three causes of death, we have considered two particular models, the (symmetric) *M*-model of Botella-Rocamora et al. (2015) and the above mentioned permuted *QsR*-model, which extends the *M*-model by permuting the rows of its **M** matrix for each municipality. Both models were implemented in `WinBUGS` and the code used to run the whole analysis (with the required complementary material) can be found as supplementary material at: `http://github.com/MigueBeneito/MatPapersMAMB-CondVsCorreg`.

We have implemented different versions of the permuted *QsR*-model in our analysis, all of them with pCAR distributions and using adjacency as criterion for defining neighboring sites. First, we considered a model with all 6=3! possible permutations of the rows of $\boldsymbol{M}$ for each spatial unit (full-*QsR* model henceforth). Nevertheless, since $\boldsymbol{M}$ rows are not identifiable for the *M*-model (Botella-Rocamora et al., 2015), we could switch any two rows of $\boldsymbol{M}$ by also switching the two corresponding columns of $\boldsymbol{\phi}$, thereby obtaining exactly the same risks. Consequently, considering all possible row permutations of $\boldsymbol{M}$ for each spatial unit could seem excessive. Therefore, in addition to the full-*QsR* model, we have also considered additional constrained permuted *QsR*-models restricting the number of permutations for the rows of $\boldsymbol{M}$ and therefore reducing their complexity. Thus, we have considered a second permuted *QsR*-model with only one possible permutation of the rows of $\boldsymbol{M}$, switching its second and third rows ((2→3)-*QsR* model henceforth). Note that the (1→2), (1→3) and (2→3)-*QsR* models would be equivalent given the unidentifiability of the columns of $\boldsymbol{\phi}$ and the rows of $\boldsymbol{M}$. All of these models basically consider the switching of two of the fitted underlying patterns in $\boldsymbol{\phi}$. Finally, we have also considered a second simplified permuted *QsR*-model, also with a single possible permutation of the rows of $\boldsymbol{M}$. For this model ((1→3→ 2)-*QsR* henceforth), the first row of **M** would be moved to the third, the third to the second, and the second to the first. Thus, this model, in contrast to (2→3)-*QsR*, switches all three rows

Table 1: Mean posterior deviance ($\bar{D}$); Effective number of parameters (p_D); Deviance Information Criterion (DIC) and logarithmic score (sum of the log-CPO values) for each of the models fitted.

Model	$\bar{D}$	p_D	DIC	log-score
M-model	6968.1	419.8	7387.9	-3776.5
Full-QsR	6962.5	433.1	7395.5	-3768.2
$(2\rightarrow3)$-QsR	6955.1	419.9	7374.9	-3759.1
$(1\rightarrow3\rightarrow2)$-$QsR$	6951.3	436.6	7387.9	-3772.7

of $\boldsymbol{M}$ instead of just two. Note that the two restricted permuted QsR-models considered, given $\boldsymbol{\phi}$ and $\boldsymbol{M}$, may reproduce 2^I different patterns in contrast to the full-QsR model, which is able to reproduce 6^I different patterns. As a result it is far more flexible and, therefore, far more parameterized.

Table 1 shows the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), and its components (mean posterior deviance and effective number of parameters) for all the models fitted. Additionally, Table 1 also reports the logarithmic score (sum of the log-CPO values) for each model (Gneiting and Raftery, 2007). Both DIC and logarithmic scores have been proposed as model selection criteria. Models with lower DICs are generally preferred, whereas for the logarithmic score models with higher values are considered as preferable. As shown in Table 1, the QsR-model outperforms the original M-model in terms of DIC, but only for the simplest $(2\rightarrow3)$-QsR model. Specifically, the $(2\rightarrow3)$-QsR and the M-model show similar complexity (pD), but the first shows a substantially better fit. In contrast, the $(1\rightarrow3\rightarrow2)$-$QsR$ and $(2\rightarrow3)$-QsR models show a similar fit but the second is more complex. Finally, the full-QsR model does not greatly improve the fit but it does visibly increase the complexity, as compared to the M-model. Regarding the logarithmic score, the M-model achieves the worst performance according to this criterion and, in contrast, the $(2\rightarrow3)$-QsR model once again shows the best results. Interestingly, according to this criterion, all three QsR-models run achieve a better performance than the original M-model.

We focus now our attention on the results of the $(2\rightarrow3)$-QsR model, the most appropriate option in terms of both DIC and log-CPOs. We have calculated the Probability Integral Transform (PIT) (Stern and Cressie, 2000; Czado et al., 2009) for each observation in our data set as a validation of the fit of the models run. PITs are used as model assessment tools since they should be uniformly distributed between 0 and 1 if the observed data has really been generated from the corresponding model. Figure 1 shows a histogram for the PITs obtained for the $(2\rightarrow3)$-QsR model. No evident departures can be appreciated on this same figure for the rest of the models fitted in this study, and thus the histograms for those models have been omitted. Figure 1 shows a distribution close to uniform on the interval [0,1], withperhaps a smaller probability for the higher values of the PITs. In any event, the Kullback-Leibler divergence for that sample of PITs versus a uniform distribution on [0,1] is equal to 0.497 units for
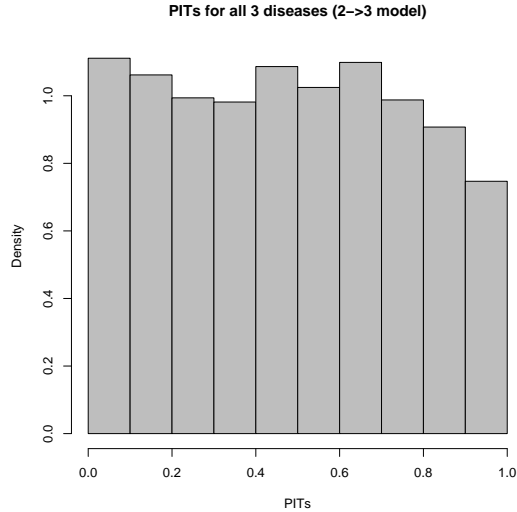
PITs for all 3 diseases (2–>3 model)

Figure 1: PITs histogram for the $(2{\rightarrow}3)$-QsR model.

the $(2{\rightarrow}3)$-QsR model, 0.512 units for the $(1{\rightarrow}3{\rightarrow}2)$-$QsR$ model, 0.513 units for the full-QsR model and 0.518 units for the M-model. Therefore, the most compatible PITs sample with a uniform distribution on [0,1] is obtained for the $(2{\rightarrow}3)$-QsR model, while, on the contrary, the largest difference with that distribution is observed for the M-model, in accordance with the model selection criteria already mentioned.

The results for the $(2{\rightarrow}3)$-QsR model show some particularly enlightening features. For example, we have calculated for Valencia (the largest city in our study) the posterior probability of its corresponding permuted $\mathbf{M}$ matrix, $\mathbf{M}_{Valencia}$, being equal to that of each of its neighbors. We have calculated this as the proportion of iterations of the MCMC in which those matrices coincided for Valencia and for each of its neighboring towns (see document Rcode.pdf in the Annex material). We have found that, for the 32 neighbors of Valencia, that posterior probability varies between 3.0% and 99.8%, and thus we have towns with a high probability of having the same $\mathbf{M}_i$ matrix as Valencia and, on the contrary, other towns where that probability is quite low.

Table 2 shows the observed cases of each disease for Valencia and those towns with $\mathbf{M}_i$ matrices that are the most and least similar to that of Valencia. Focusing on cirrhosis and lung cancer (the causes with the highest mortality rates), we see that the ratio of those mortalities for Valencia is 2.92(=7938/2722). For the municipalities with a similar $\mathbf{M}_i$ matrix to that of Valencia the ratio is 2.65 and for the towns with the most different $\mathbf{M}_i$ matrices the ratio is 5.34, which is very different to that of Valencia. Note that this same relationship also holds

Table 2: Observed cases of mortality for each cause and group of towns and cities: Valencia, towns with the most similar **M** to that of Valencia and towns with the least similar **M** matrix, respectively.

Region	Cirrhosis	Lung Cancer	Oral Cancer
Valencia	2722	7938	734
Neighbors i with $P(\mathbf{M}_{Valencia} = \mathbf{M}_i) > 0.8$	542	1439	126
Neighbors i with $P(\mathbf{M}_{Valencia} = \mathbf{M}_i) < 0.2$	120	641	65

for oral cancer and cirrhosis but, conversely, the ratio between lung and oral cancer is approximately constant for the three groups considered. Thus, the permuted (2→3)-QsR-model seems to permute the **M** matrix for those pairs of neighbors whose data show a markedly different performance for cirrhosis in comparison to the tumor-related causes of death. This permutation allows neighbors with a different performance in terms of cirrhosis to be weighted in a different manner to the rest of the neighbors. Interestingly, the log-CPOs for the (2→3)-QsR-model (see the supplementary `Rcode.pdf` document) show that this model achieves the largest benefit for this disease, in comparison to the M-model, while that improvement for the remaining diseases is much milder.

Table 3 shows the cross-covariance matrices (their posterior means) fitted for Valencia-Sollana and Valencia-Moncada, respectively. Sollana and Moncada are Valencia's neighbors with the lowest and highest posterior probabilities, respectively, of their $\mathbf{M}_i$ matrices to be equal to that of Valencia. On the one hand, we see that the cross-covariance matrix is symmetric for Valencia-Moncada, up to 3 digits, but not for Valencia-Sollana. This illustrates how having different $\mathbf{M}_i$ matrices makes the corresponding cross-covariance matrix for two municipalities asymmetric. Moreover, that cross-covariance matrix will be different for any two pairs of neighbors depending on the resemblance of their data, in contrast to other asymmetrical modeling proposals in the literature. On the other hand, important differences can be noticed between both two cross-covariance matrices in Table 3, particularly for their first diagonal cells, those in charge of controlling the covariance for cirrhosis mortality for this pair of sites. Thus, we see that the value of this cell is high between Valencia and Moncada but for Valencia-Sollana that covariance is much milder, even negative. Therefore, the permuted QsR-model detects that, for some neighboring municipalities (but only some), the risk of cirrhosis could be uncorrelated between them and flips the **M** matrix for some of these municipalities in order to reproduce that performance. This different contribution of each neighbor is a performance that cannot be reproduced by symmetric (traditional) M-models.

We wish to conclude this section by discussing the computational performance of the proposed models. The computing times for the models run in this section were the following: M-model 10.7 minutes; full-QsR model 14.0 minutes; (2→3)-QsR model 12.1 minutes; (1→3→ 2)-QsR model 12.3 minutes. It can be appreciated that the QsR models increase computing times, as compared to the M-model, in a reasonable amount of time. The highest increase in comput-

Table 3: Cross-covariance matrices for Valencia-Sollana and Valencia-Moncada. Sollana and Moncada are Valencia's neighbors with the lowest and highest probabilities, respectively, of having the same **M** matrix as Valencia. Causes of death are shown in the following order: cirrhosis, lung and oral cancer.

Cross-Cov. Valencia-Sollana			Cross-Cov. Valencia-Moncada		
-0.050	0.112	0.106	0.229	0.122	0.163
0.089	0.154	0.162	0.122	0.156	0.171
0.103	0.167	0.244	0.163	0.171	0.261

ing time is achieved for the most complex option (full-QsR model) with a 31% increase over the original M-model computing time. In addition, we have run the same models as in the previous analysis but now for a multivariate analysis of 5 diseases. We skip most of the details in this analysis, such as the diseases studied, since this is carried out just for computational goals. The computing times for this new study were the following: M-model 24.7 minutes; full-QsR model 121.1 minutes; $(2\to3)$-QsR model 37.6 minutes; $(1\to3\to 2)$-QsR model 42.4 minutes. According to these results, the $(2\to3)$-QsR and $(1\to3\to 2)$-QsR models have a moderately worse performance in comparison to the M-model, as compared to the multivariate analysis of 3 diseases. Thus, for the original analysis with 3 diseases these QsR models spent around 15% more time than the corresponding M-model, while for the 5-disease analysis the additional computing time requirements increase until by 82%. However, much more evident differences are noticed for the full-QsR model in this new analysis, which is 31% more expensive (in computing time terms) than the M-model for the 3-disease analysis but 490% more expensive for the 5-disease analysis. Thus, the full-QsR model scales poorly as a function of the number of diseases considered in the multivariate study. This seems a consequence of the number of possible M_i matrices for this model, which increases from $6(=3!)$ in the 3-disease analysis to $120(=5!)$ in the 5-disease analysis. For the other 2 QsR-models considered, with much better performance, this quantity remains constant for both studies considered.

5 Conclusions

The main goal of this paper has been to establish some relationships between the classes of (multivariate and univariate) conditional and coregionalized models. This paper makes some contributions in that direction that may help to put some order in those three classes and provide an overall insight into the different main approaches followed to develop MGMRF proposals. These links have both theoretical and practical consequences. On the one hand, for any model, having an alternative formulation as a model of the rest of the classes may provide additional insight regarding sensible modifications that could make that proposal even more suitable for some particular datasets. Moreover, seeing that

model as an element of the other classes will also make the theoretical results of that second class available to the original model. On the other hand, the links between these classes can also yield important advantages in practical (computational) terms. In our experience, coregionalization models are a far more advantageous alternative to conditional proposals in practical terms. For instance, M-models have been used for fitting multivariate studies of tens of diseases and their use avoids the need to deal with Kronecker products or some other non-standard matrix algebra tools which may not be implemented in regular Bayesian software packages. The results in this paper provide some clues on how to formulate some conditional models as coregionalization models. As mentioned, this reformulation may provide important computational benefits.

As pointed out in Section 3, the results in this paper show that some models already proposed in the literature may be overparameterized. As shown, any multivariate conditional model can be expressed as an equivalent model of that same class with diagonal precision matrices. In that case, apparently, there is not any theoretical reason for considering general symmetric positive definite matrices for $\mathbf{\Omega}_i$, $i = 1, ..., I$, unless the set of autoregression matrices $\{\boldsymbol{\beta}_{ii'} : i, i' = 1, ..., I\}$ is heavily constrained. Otherwise, both precision and autoregression matrices may play the same role in the model, leading to both terms being unidentifiable. This result suggests caution is in order when proposing conditional models, that allow extreme flexibility, since that flexibility could make some terms in the model unidentifiable. This cautiousness should be even more extreme when proposing hybrid models merging all three approaches introduced in this paper, such as for example non-pure coregionalization models. In those cases, the joint use of several tools for inducing dependence could make it even more likely that several components played similar (or plainly the same) roles, making identifiability problems even more significant.

Proposing QsR-models is the second main contribution of this paper. Although it was not the principal aim of this work, that proposal is also of substantial interest and opens coregionalization models to the modeling of MGMRF with (order-free) asymmetric cross-covariances, in contrast to what was previously believed. In our opinion, this is a second important theoretical contribution of this paper. Nevertheless, besides the theoretical interest of QsR-models in general, permuted QsR-models have shown themselves to be interesting simplifications of the original QsR-models that, despite their theoretical interest, are clearly over-parameterized and therefore of limited use in practical terms. The permuted counterpart, in contrast, has proven to be a computationally affordable option with particularly good results for its more parsimonious versions. Nevertheless, more experience is needed with this model in order to propose and explore permuted QsR-models that are flexible enough, and at the same time avoid the potential overparameterization of the full QsR-model. These models could provide classes of flexible and computationally affordable models of evident practical interest. The real potential of QsR-models has not been explored in depth, as the main aim of this paper was to provide a thorough unifying view of coregionalized and conditional MGMRFs. Nevertheless, our preliminary re-

sults with the *QsR*-models suggest that it is surely worth conducting a deeper study of that approach in order to assess its real potential in practice. The current results in that direction seem quite promising.

Acknowledgements

Annex

Theorem 1 *Let $\boldsymbol{\zeta}$ be an MGMRF in $\mathcal{C}_M$ defined as:*

$$\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}\boldsymbol{\zeta}_{i'\cdot}^T, \boldsymbol{\Sigma}_i^{-1}), i = 1,...,I,$$

then $\boldsymbol{\zeta}$ also belongs to $\mathcal{C}_U$. In particular:

$$\zeta_{ij}|vec(\boldsymbol{\zeta}_{-(ij)}) \sim N\left(\sum_{i'\sim i}\beta^*_{iji'j}\zeta_{i'j} + \sum_{j'\neq j}\beta^*_{ijij'}\zeta_{ij'} + \sum_{j'\neq j}\sum_{i'\sim i}\beta^*_{iji'j'}\zeta_{i'j'}, \sigma_{ij}^{-2}\right)$$

for $\beta^*_{iji'j'} = (diag(\boldsymbol{\Sigma}_i^{-1})^{-1}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_{ii'})_{jj'}$ and $\sigma_{ij}^{-2} = (\boldsymbol{\Sigma}_i^{-1})_{jj}$.

Proof:

Elemental properties of the multivariate Normal distribution (see for example Appendix A of Gelman et al. (2014)) state that if $\boldsymbol{\theta} \sim N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\theta_j|\boldsymbol{\theta}_{-j} \sim N(m_j, \sigma_j^2)$ for:

$$m_j = \mu_j - \sum_{l\neq j}\frac{(\boldsymbol{\Sigma}^{-1})_{jl}}{(\boldsymbol{\Sigma}^{-1})_{jj}}(\theta_l - \mu_l)$$

and $\sigma_j^{-2} = (\boldsymbol{\Sigma}^{-1})_{jj}$.

If we applied this result to $\boldsymbol{\zeta}_{i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}\boldsymbol{\zeta}_{i'\cdot}^T, \boldsymbol{\Sigma}_i)$, we would obtain that $\zeta_{ij}|\boldsymbol{\zeta}_{-(ij)} = \zeta_{ij}|\boldsymbol{\zeta}_{i(-j)}, vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N\left(m_{ij}, \sigma_{ij}^2\right)$ for $\sigma_{ij}^{-2} = (\boldsymbol{\Sigma}_i^{-1})_{jj}$

and:

$$
\begin{aligned}
m_{ij} &= \sum_{i'\sim i}(\boldsymbol{\beta}_{ii'})_j.\boldsymbol{\zeta}_{i'.}^T - \sum_{l\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\zeta_{il} - \sum_{i'\sim i}(\boldsymbol{\beta}_{ii'})_l.\boldsymbol{\zeta}_{i'.}^T) \\
&= \sum_{i'\sim i}\left((\boldsymbol{\beta}_{ii'})_{jj}\zeta_{i'j} + \sum_{j'\neq j}(\boldsymbol{\beta}_{ii'})_{jj'}\zeta_{i'j'}\right) - \sum_{l\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}\zeta_{il} + \\
&\quad + \sum_{l\neq j}\sum_{i'\sim i}\left(\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj}\zeta_{i'j} + \sum_{j'\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj'}\zeta_{i'j'}\right) \\
&= \sum_{i'\sim i}\left((\boldsymbol{\beta}_{ii'})_{jj} + \sum_{l\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj}\right)\zeta_{i'j} - \sum_{j'\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jj'}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}\zeta_{ij'} + \\
&\quad + \sum_{j'\neq j}\sum_{i'\sim i}\left((\boldsymbol{\beta}_{ii'})_{jj'} + \sum_{l\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj'}\right)\zeta_{i'j'}.
\end{aligned}
$$

defining $\beta_{iji'j'}^*$ as:

$$
(\boldsymbol{\beta}_{ii'})_{jj'}+\sum_{l\neq j}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj'} = \sum_{l=1}^{J}\frac{(\boldsymbol{\Sigma}_i^{-1})_{jl}}{(\boldsymbol{\Sigma}_i^{-1})_{jj}}(\boldsymbol{\beta}_{ii'})_{lj'} = (diag(\boldsymbol{\Sigma}_i^{-1})^{-1}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\beta}_{ii'})_{jj'}
$$

and bearing in mind that $\boldsymbol{\beta}_{ii} = -\boldsymbol{I}_J$, the latter expression of m_{ij} can be expressed as:

$$
\sum_{i'\sim i}\beta_{iji'j}^*\zeta_{i'j} + \sum_{j'\neq j}\beta_{ijij'}^*\zeta_{ij'} + \sum_{j'\neq j}\sum_{i'\sim i}\beta_{iji'j'}^*\zeta_{i'j'}
$$

as we wanted to prove.

Theorem 2 *Given $\boldsymbol{\zeta}$ an MGMRF in $\mathcal{C}_U$, $\boldsymbol{\zeta}$ can also be formulated as a multivariate conditional model in $\mathcal{C}_M$. Specifically, if:*

$$
\zeta_{ij}|vec(\boldsymbol{\zeta}_{-(ij)}) \sim N\left(\sum_{i'\sim i}\beta_{iji'j}\zeta_{i'j} + \sum_{j'\neq j}\beta_{ijij'}\zeta_{ij'} + \sum_{j'\neq j}\sum_{i'\sim i}\beta_{iji'j'}\zeta_{i'j'}, \sigma_{ij}^{-2}\right)
$$

then it can be also equivalently expressed as:

$$
\boldsymbol{\zeta}_{-i.}|vec(\boldsymbol{\zeta}_{-i.}^T) \sim N_J\left(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}^*\boldsymbol{\zeta}_{i'.}^T, diag(\boldsymbol{\sigma}_{i.}^{-2})\right), \ i = 1,...,I.
$$

for $\boldsymbol{\beta}_{ii'}^* = (\beta_{iji'j'})_{j,j'=1}^J$.

Proof:

Given the array $\boldsymbol{\beta} = (\beta_{iji'j'})$ and $\boldsymbol{\sigma} = (\sigma_{ij})$ in the univariate conditional model

above, define $\boldsymbol{\Sigma}_i = diag(\boldsymbol{\sigma}_{i\cdot}^2)$ and $\boldsymbol{\beta}_{ii'}^* = (\beta_{iji'j'})_{j,j'=1}^J$. Let us consider the following multivariate conditional model:

$$\zeta_{-i\cdot}|vec(\boldsymbol{\zeta}_{-i\cdot}^T) \sim N_J\left(\sum_{i'\sim i}\boldsymbol{\beta}_{ii'}^*\boldsymbol{\zeta}_{i'\cdot}^T, diag(\boldsymbol{\sigma}_{i\cdot}^2)\right), \ i = 1,...,I.$$

Result 1, when applied to this multivariate conditional model, yields the univariate model given in the formulation of this result, which shows the equivalence of both formulations.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36:192–236.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Hiu-Wan, K., and Silkey, M. (1997). Natural variability of benthic species composition in the Delaware Bay. *Environmental and Ecological Statistics*, 5:95–115.

Botella-Rocamora, P., López-Quílez, A., and Martinez-Beneito, M. A. (2012). Spatial moving average risk smoothing. *Statistics in Medicine*, 32:2595–2612.

Botella-Rocamora, P., Martinez-Beneito, M. A., and Banerjee, S. (2015). A unifying modeling framework for highly multivariate disease mapping. *Statistics in Medicine*, 34(9):1548–1559.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*, pages 45–64. Oxford University Press.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–25.

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.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Greco, F. P. and Trivisano, C. (2009). A multivariate CAR model for improving the estimation of relative risks. *Stat. Med.*, 28(12):1707–1724.

Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics*, 7:65–81.

Ippoliti, L., Martin, R. J., and Romagnoli, L. (2018). Efficient likelihood computations for some multivariate Gaussian Markov random fields. *Journal of Multivariate Analysis*, 168:185–200.

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

Kim, H., Sun, D., and Tsutakawa, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *J. Amer. Statist. Assoc.*, 96(456):1506–1521.

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

MacNab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: a general framework for coregionalized multivariate CAR models. *Statistics in Medicine*, pages 3827–3850.

MacNab, Y. C. (2018). Some recent work on multivariate Gaussian Markov random fields. *TEST*, 27(3):497–541.

Mardia, K. V. (1988). Multidimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivariate Anal.*, 24(2):265–284.

Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.

Martinez-Beneito, M. A. (2018). Comments on: Some recent work on multivariate Gaussian Markov random fields. *TEST*, 27(3):542–544.

Martinez-Beneito, M. A. and Botella Rocamora, P. (2019). *Disease mapping from foundations to multidimensional modeling.* CRC Press.

Martinez-Beneito, M. A., Botella-Rocamora, P., and Banerjee, S. (2017). Towards a multidimensional approach to Bayesian disease mapping. *Bayesian Analysis*, 12:239–259.

Sain, S. and Cressie, N. (2007). A spatial analysis of multivariate lattice data. *Journal of Econometrics*, 140:226–259.

Sain, S., Furrer, R., and Cressie, N. (2011). A spatial analysis of multivariate output from regional climate models. *The Annals of Applied Statistics*, 5(1):150–175.

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.

Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, 19:2377–2397.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer.