



# Rejoinder on: Some recent work on multivariate Gaussian Markov random fields

Ying C. MacNab<sup>1</sup>

Published online: 12 September 2018

© Sociedad de Estadística e Investigación Operativa 2018

## Abstract

I thank the discussants, Miguel A. Martínez-Beneito, Fedel Greco, Carlo Trivisano, Stephan R Sain, and Reinhard Furrer, for their insightful and stimulating commentary. The rejoinder is organized in five sections: (1) the M-based models, (2) posterior sensitivity to prior choices for  $C$  and  $\Sigma$ , (3) stationary and non-stationary (M)GMRFs, (4) various approaches to model formulation and related applications, and (5) statistical computation.

## 1 The M-based models

Martínez-Beneito highlights two main advantages of coregionalization models: computational convenience and validity by construction. In particular, he emphasizes the computational advantages of the M-model construction. Indeed, hierarchically formulated M-models and associated Bayesian computational methods are shown to be computationally efficient in handling multivariate and multiarray spatial lattice data of many variables.

The M-model proposal has two issues: identification and interpretation. It seems to me that a good understanding of the two issues is important for both the methodological and practical reasons. Here, I briefly discuss these issues and the need for additional research.

Consider any  $p$ -variate M-model with  $p$  variable-specific spatial dependence parameters, denoted M-model  $(c_1, c_2, \dots, c_p)$  hereafter. It is mentioned in Botella-Rocamora et al. (2015), MacNab (2016b), and the present paper that the spatial parameters  $c_1, c_2, \dots, c_p$  and the M-matrix therein are not identifiable. It seems to me that the gain in computational efficiency for M-models comes at a price: the data loss

---

This rejoinder refers to the comments available at <https://doi.org/10.1007/s11749-018-0606-2>; <https://doi.org/10.1007/s11749-018-0607-1>; <https://doi.org/10.1007/s11749-018-0609-z>.

---

✉ Ying C. MacNab  
ying.macnab@ubc.ca

<sup>1</sup> School of Population and Public Health, University of British Columbia, Vancouver, Canada

**Table 1** DIC results for indicated models. The Minnesota cancer mortality data and the BC adverse medical events data

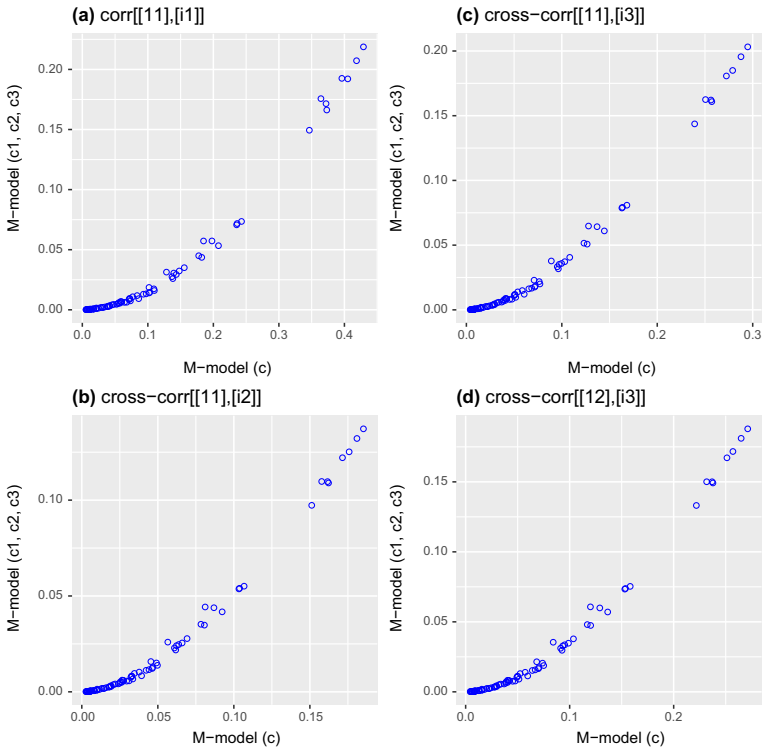
Model	Dbar	$pD$	DIC
M-model ( $c_1, c_2, c_3$ ), $\mathbf{M}[i, j] \sim \text{dflat}()$	1422	78	1501
M-model ( $c$ ), $\mathbf{M}[i, j] \sim \text{dflat}()$	1424	69	1493
Separable MpCAR, $\mathbf{\Gamma} \sim \text{Wishart}(\text{diag}(1), 3)$	1408	99	1507
Separable MpCAR, $\mathbf{\Gamma} \sim \text{Wishart}(\text{diag}(0.1), 3)$	1434	67	1501
MiCAR, $\mathbf{\Gamma} \sim \text{Wishart}(\text{diag}(1), 3)$	1411	99	1510
MiCAR, $\mathbf{\Gamma} \sim \text{Wishart}(\text{diag}(0.1), 3)$	1430	68	1498
Bspline, 4 knots, M-model ( $c_1, c_2, \dots, c_8$ ), $\mathbf{M}[i, j] \sim \text{dflat}()$	3931	339	4270
Bspline, 4 knots, M-model ( $c$ ), $\mathbf{M}[i, j] \sim \text{dflat}()$	3932	338	4270

control and identification of the spatial dependence parameters. An important question arises: What are the additional benefit(s) the M-models might bring, compared to their counterparts of separable models? Examples of competing separable models include the M-model ( $c$ ) (with a general spatial parameter  $c$ ), the separable models of the Mardia family, and the intrinsic multivariate CAR (MiCAR) models, to name a few; see Table 1 for a brief illustrative DIC comparison. Separable models also have similar or greater computational advantages. Additionally, data (partially) inform on the general spatial parameter in a separable model.

Briefly illustrated here using the two data sets presented in the present paper, the results of my recent study suggest that (i) the Markov chain Monte Carlo (MCMC) implementation for the M-model ( $c$ ) may be more stable, (ii) the spatial parameter in M-model ( $c$ ) is identifiable, (iii) the M-model ( $c$ ) may outperform M-model ( $c_1, c_2, \dots, c_p$ ) in terms of DIC (see Table 1), (iv) the M-model ( $c$ ) may lead to less posterior shrinkage of the marginal correlation and cross-correlation functions (see Fig. 1), and (v) the two models can produce nearly identical posterior relative risk smoothing, prediction, and inference (see Fig. 2).

In Botella-Rocamora et al. (2015), the posterior estimates of the within-location covariance matrix  $\mathbf{\Sigma}$  are used to draw inference for (log) relative risk correlations between diseases. Due in part to the complex “entanglement” of the spatial and non-spatial parameters in the M-model and in MGMRFs in general, and perhaps due in part to the area-specific scaling factors, such interpretation of  $\mathbf{\Sigma}$  for inference on the pair-wise associations between diseases should be questioned. For example, it seems to me that there is a tendency for the posterior estimates of the correlation parameters in  $\mathbf{\Sigma} (= \mathbf{M}\mathbf{M}^\top)$  to overestimate the disease risk associations. For the Minnesota cancer mortality data, the Pearson correlation coefficient (PCC) for the esophageal and lung cancers is 0.28. But, the posterior median (and standard deviation) of the associated correlation parameter in  $\mathbf{\Sigma}$  is  $r_{13} = 0.66$  (0.22) for the estimated M-model A and  $r_{13} = 0.73$  (0.21) for the estimated M-model B. Table 2 presents the estimated non-spatial correlation parameters for the estimated M-model and the MiCAR, respectively. The correlation estimates for the MiCAR A are the closest to the PCCs.

These results on the M-models warrant a further investigation. For example, assessing the M-models via a simulation study may offer insights into their utility as spatial smoothers in multivariate disease mapping. The use of the M-models (with variable-

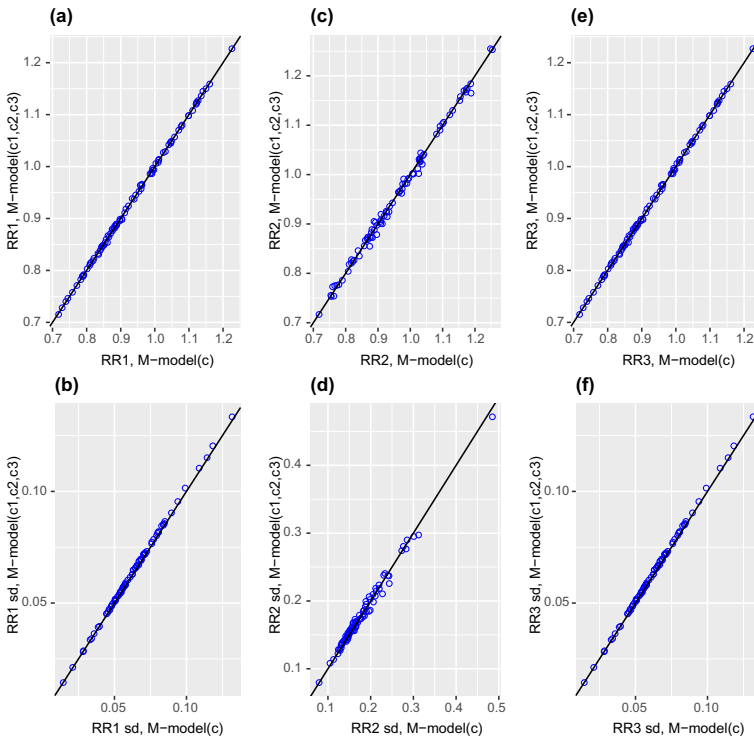


**Fig. 1** Illustrative comparisons of the posterior estimates of correlations (corr) and cross-correlations (cross-corr) functions between the estimated M-model ( $c_1, c_2, c_3$ ) and M-model ( $c$ ). The estimates are illustrated using the results for county 1, which has 8 first-order neighbors. Each of the plots shows a cluster of higher estimates of correlations and cross-correlations between the county 1 and its 8 first-order neighbors, compared to the estimates between the county 1 and its higher-order neighbors. The Minnesota cancer mortality data

specific spatial parameters), rather than the counterparts of separable models, for modeling multi-way data, as presented in Martinez-Beneito et al. (2017), also warrants practically useful motivation, interpretation, and numeric assessment and comparison.

## 2 Posterior sensitivity to prior choices for $C$ and $\Sigma$

In the context of multivariate disease mapping where data may contain limited information, Greco and Trivisano raise important practical issues concerning hyperprior choices for MCARs and effect on model comparison and selection. They point out that (i) posterior sensitivity to prior specifications for  $C$  and  $\Sigma$  may have a complex impact on model selection (using DIC) and (ii) the currently used prior specifications over the reparameterizations of the matrices  $C_s$ ,  $C$ , and  $\Sigma$  may lead to order-sensitive posterior risk shrinkage, predictions, and inference. In disease mapping and small area estimation, notable posterior sensitivities to hyperprior specifications of (M)GMRFs are not uncommon and should be reported as important indications of statistical uncertainty.



**Fig. 2** Illustrative comparisons of posterior estimates, median and standard deviation (sd), of relative risks between the estimated M-model (c) and M-model (c<sub>1</sub>, c<sub>2</sub>, c<sub>3</sub>). The Minnesota cancer mortality data

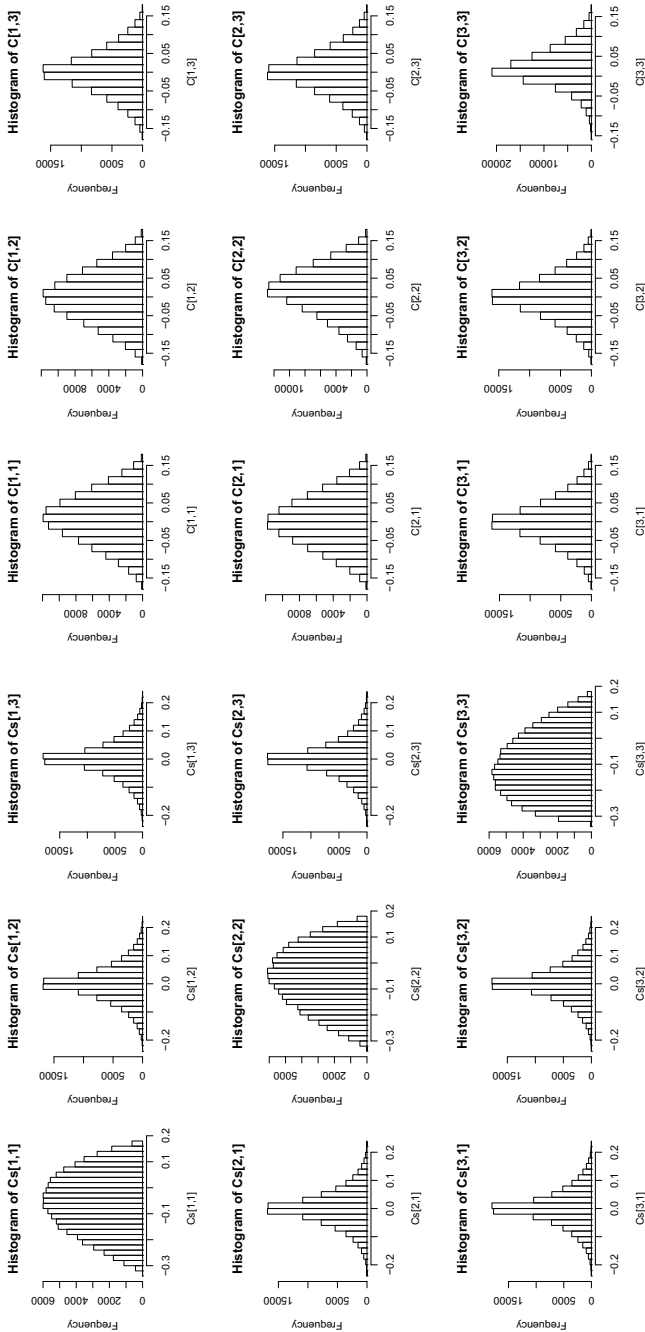
**Table 2** Posterior median and standard deviation (sd) of correlation parameters in  $\Sigma$ , derived from the estimated M-models and MiCAR

Parameter	M-model A		M-model B		MiCAR A		MiCAR B		PCC
	Median	SD	Median	SD	Median	SD	Median	SD	
$\rho_{12}$	0.49	0.43	0.58	0.38	0.16	0.25	0.50	0.31	0.06
$\rho_{13}$	0.66	0.22	0.73	0.21	0.25	0.16	0.59	0.17	0.28
$\rho_{23}$	0.62	0.29	0.71	0.27	0.15	0.18	0.53	0.24	-0.06

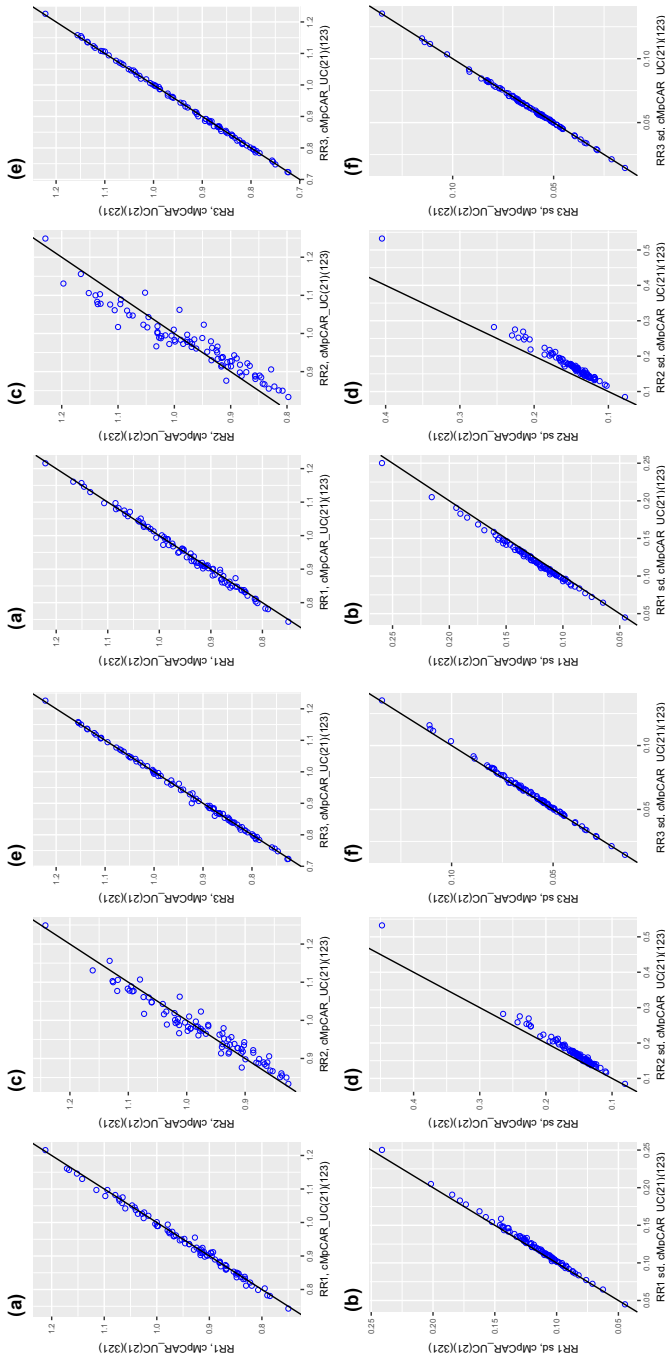
M-model A: the M-model with prior  $M_{jl} \sim \text{dflat}()$ ; M-model B: M-model with prior  $M_{jl} \sim N(0, \sigma^{-2}), \sigma \sim \text{Unif}(0, 10)$ ; MiCAR A: MiCAR with  $\Gamma \sim \text{Wishart}(\text{diag}(1), 3)$ ; MiCAR B: MiCAR with  $\Gamma \sim \text{Wishart}(\text{diag}(0.1), 3)$ ; PPC: Pearson correlation coefficients based on the crude SMRs. The Minnesota cancer mortality data

In what follows, I further explain the issues raised by Greco and Trivisano with illustrative results of additional multivariate analysis of the Minnesota cancer data.

I begin with a brief illustration and explanation of how and why the uniform priors placed on the eigenvalues of a symmetric matrix  $C_s$ , or on the singular values of an asymmetric  $C$ , with uniform priors on the associated Givens angles, might impose *a priori* restrictions on the elements of  $C_s$  or  $C$ . Figure 3 presents the resulting element-



**Fig. 3** Illustration of shrinkage priors on the spatial parameters, resulting from placing the PDC on the (descending) ordered eigenvalues of  $C_s$  (the 9 plots on the left) and the PDC on the (descending) ordered singular values of  $C$  (the 9 plots on the right), respectively. The PDC ranges are  $(-0.322, 0.178)$  and  $(0, 0.178)$ , respectively, calculated from the eigenvalues of  $W$  of connected first-order neighbors in 87 Minnesota counties of USA



**Fig. 4** Posterior estimates (median and standard deviation) of relative risks for the linear coregionalization cMpCAR<sub>UC</sub>(21). Positive definite constraint is placed on the singular values of  $C$  via singular value decomposition. The coregionalization coefficients matrix  $A = P(\theta)eP(\theta)^T$  is the square root of  $\Sigma$ . The Minnesota cancer data

wise prior distributions for all elements of  $C_s$  and  $C$ , respectively. These element-wise histograms were calculated based on 10,000 samples of  $C_s = P(\theta)eP(\theta)$  or  $C = P(\theta_L)sP(\theta_R)$ , where the ordered eigenvalues  $e$  were simulated from  $\text{Unif}(-0.322, 0.178)$ , the ordered singular values  $s$  from  $\text{Unif}(0, 0.178)$ , and the Givens angles  $\theta$ ,  $\theta_L$ , and  $\theta_R$  from  $\text{Unif}(-\pi/2, \pi/2)$ .

As noted by Greco and Trivisano, *element-wise* prior patterns can be observed from Fig. 3. Placing the above-mentioned priors on the eigenvalue decomposition of  $C_s$  leads to skewed prior distributions on the diagonal elements of  $C_s$ . Heavier prior restrictions toward zero are placed on the off-diagonal elements of  $C_s$ . The skewed prior distributions, from right-skewed to the left-skewed (see the 9 plots on the left), correspond to the descending order of the eigenvalues from 0.178 down to  $-0.322$ . Likewise, the patterns of increasing prior concentration toward zero, over the diagonal and the off-diagonal elements of  $C$  (illustrated in the 9 plots on the right), are also in line with the descending order of the positive singular values, from the upper limit 0.178 down toward 0.

It should be mentioned that these patterns are not due to the use of priors on the Givens angles but are the result of eigen- or singular value decomposition with *ordered* eigen- or singular values for unique decomposition of  $C_s$  or  $C$ . It is readily verified that these patterns should disappear if the priors for  $C_s$  or  $C$  were simulated from the same reparameterization but with un-sorted eigen- or singular values. Notice that the descending or ascending ordering of the eigen- or singular values is necessary to enable identification of  $C_s$  or  $C$  via its *unique* decomposition. Figure 3 also indicates that, compared to placing priors on the eigenvalue decomposition of  $C_s$ , placing priors on the singular value decomposition of  $C$  may lead to greater posterior shrinkage on the diagonal elements  $\{c_{jj}, \forall j\}$  but less posterior shrinkage on the off-diagonal elements  $\{c_{jl}, \forall j \neq l\}$ .

In the Minnesota cancer mapping application, and for the cMpCARs of the Type II decompositions, notable order sensitivities were observed from the resulting deviance information measures (see Table 3) and from the posterior estimates of spatial and non-spatial parameters (see Table 4). The posterior estimates of relative risks were relatively unchanged for esophageal and lung cancers, respectively, with modest order sensitivity for laryngeal cancer (see Fig. 4). Similar results are also observed from the cMpCARs of the Type I decomposition (MacNab 2018).

Placing hierarchical priors on the elements of  $C$ , as presented in MacNab (2016b, 2018), may be one approach to order-invariant estimation of  $C$  and MGMRFs. My recent case studies seem to indicate that placing hierarchical priors on the elements

**Table 3** DIC results for indicated models

Model	Dbar	pD	DIC
cMpCAR <sub>UC</sub> (C) <sup>†</sup> , $y_1, y_2, y_3$	1423	72	1495
cMpCAR <sub>UC</sub> (C) <sup>†</sup> , $y_3, y_2, y_1$	1427	68	1494
cMpCAR <sub>UC</sub> (C) <sup>†</sup> , $y_2, y_3, y_1$	1424	56	1480

The Minnesota cancer mortality data

†: Priors are placed on singular value decomposition of  $C$

**Table 4** Posterior estimates for the linear coregionalization  $\text{cMpCAR}_{\text{UC}}(\mathbf{C}, \mathbf{A})(21)$ , with constraint  $s_j \in (0, c_{\max}), \forall j$ , on the singular value decomposition of an asymmetric matrix of  $\mathbf{C}$ ,  $\mathbf{A} = P(\boldsymbol{\theta})\mathbf{e}^s P(\boldsymbol{\theta})$ , and for the three ways of ordering the variables indicated in table

Node	$(\xi_1, \xi_2, \xi_3)$		$(\xi_3, \xi_2, \xi_1)$		$(\xi_2, \xi_3, \xi_1)$	
	Median	SD	Median	SD	Median	SD
$c_{11}$	0.15	0.36	0.46	0.41	0.42	0.46
$c_{22}$	0.10	0.41	0.37	0.33	0.38	0.32
$c_{33}$	0.67	0.22	0.67	0.22	0.61	0.20
$c_{12}$	0.04	0.39	0.01	0.39	0.09	0.39
$c_{21}$	0.00	0.37	0.01	0.38	0.07	0.39
$c_{13}$	0.28	0.28	0.20	0.30	0.11	0.32
$c_{31}$	0.37	0.27	0.31	0.29	0.23	0.29
$c_{23}$	0.17	0.30	0.15	0.29	0.18	0.28
$c_{32}$	0.24	0.28	0.28	0.29	0.31	0.28
$\rho_{12}$	0.31	0.29	0.16	0.23	0.30	0.30
$\rho_{13}$	0.43	0.22	0.44	0.23	0.37	0.19
$\rho_{23}$	0.22	0.21	0.36	0.27	0.44	0.28
$\sigma_1$	0.25	0.09	0.24	0.08	0.26	0.09
$\sigma_2$	0.31	0.13	0.27	0.12	0.24	0.12
$\sigma_3$	0.19	0.03	0.18	0.03	0.19	0.03

The Minnesota cancer mortality data

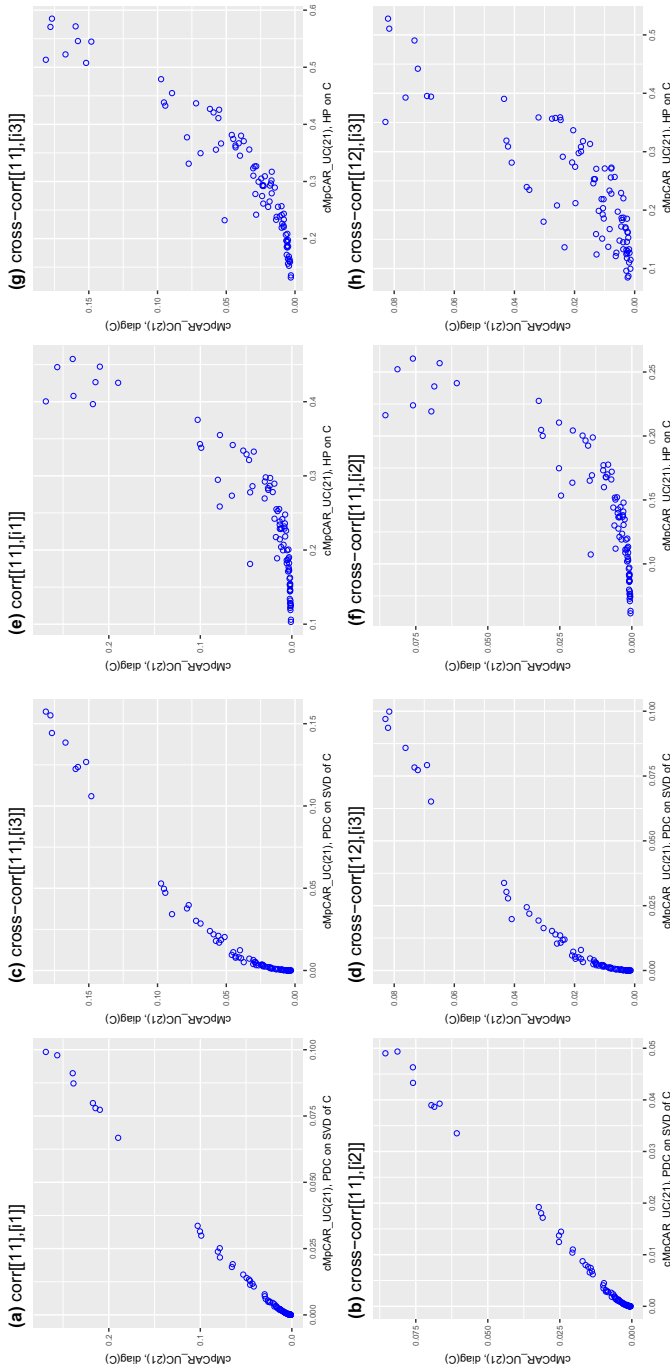
of  $\mathbf{C}$  may impose less posterior shrinkage to  $\mathbf{C}$ , perhaps less shrinkage to the diagonal elements of  $\mathbf{C}$  (MacNab 2016b, 2018). The plots (a)–(d) in Fig. 5 illustrate that the estimated correlation and cross-correlation functions of the  $\text{cMpCAR}_{\text{UC}}(\mathbf{C}, \mathbf{A})$  with positive definiteness constraint (PDC) and associated priors on the singular value decomposition (SVD) of  $\mathbf{C}$  are similar to, but overall lower than, those of the  $\text{cMpCAR}_{\text{UC}}(\mathbf{c}, \mathbf{A})$ , the MGMRF with a diagonal matrix  $\mathbf{C} = \text{diag}(\mathbf{c})$ . Note that the PDC and associated priors on the SVD of  $\mathbf{C}$  may impose considerable shrinkage to both the diagonal and off-diagonal elements of  $\mathbf{C}$ , which might be a reason that the estimated correlation and cross-correlation functions of the  $\text{cMpCAR}_{\text{UC}}(\mathbf{C}, \mathbf{A})$ , in Fig. 5, plots (a)–(d), are overall lower than those of the  $\text{cMpCAR}_{\text{UC}}(\mathbf{c}, \mathbf{A})$ . In contrast, the plots (e)–(h) in Fig. 5 seem to suggest that element-wise HPs on  $\mathbf{C}$  may lead to notably less posterior shrinkage to the correlation and cross-correlation functions.

Greco and Trivisano also comment and illustrate the impact of sensitivity to prior specification for  $\boldsymbol{\Sigma}$  on model comparison and selection. I agree with them that in the present paper the observed differences between the models may be influenced by the different prior specifications for  $\boldsymbol{\Sigma}$  or  $\boldsymbol{\Gamma}$ . Briefly illustrated in Table 1, when data contain limited information, posterior sensitivity can be observed from the same (and relatively simple) model but different (Wishart) prior specifications for  $\boldsymbol{\Gamma}$ .

### 3 Stationary and non-stationary (M)GMRFs

Sain and Furrer comment that Markov random fields do not, in general, lead to stationary models. This would be true for the coregionalization models. In general, the





**Fig. 5** Illustrative comparisons of posterior estimates of correlations (corr) and cross-correlations (cross-corr) between the estimated  $cMpCAR_{UC}(C, A)$  and  $cMpCAR_{UC}(diag(c_1, c_2, c_3), A)$ , where  $A$  is the square root of  $\Sigma$ . The estimates are presented for county 1. The matrix  $C$  in  $cMpCAR_{UC}(C, A)$  is allowed to be asymmetric. The estimated correlations and cross-correlations are based on (i) placing positive definiteness constraint (PDC) priors on the singular value decomposition (SVD) of  $C$ , the 4 plots on the left, or (ii) placing hierarchical priors (HP) on the elements of  $C$ , the 4 plots on the right. The Minnesota cancer mortality data

so-called edge effects lead to latent (M)GMRFs with marginal correlations that differ by location. While not discussed in the present paper, formulations of stationary (M)GMRFs for rectangular lattice-neighborhood schemes (with appropriate boundary conditions/adjustments) are discussed in Besag (1972, 1974) and Mardia (1988). Similar approaches can be taken to formulate stationary latent fields that lead to stationary coregionalization models. As mentioned by Sain and Furrer, stationary (M)GMRFs may be motivated and formulated by problem-driven considerations of neighborhood structures. Compared to non-stationary (M)GMRFs, these models typically involve smaller number of unknown parameters and often have computational advantages, say, in terms of scalability and efficiency.

In the present paper, some non-stationary (M)GMRFs with locally varying (adaptive) spatial and/or scale parameters are briefly outlined. These models are indeed complex and contain many parameters. Briefly mentioned in the paper, locally adaptive (M)GMRFs may be considered for their flexibility of modeling complex multivariate interaction and dependence structures, perhaps facilitated by additional data for covariates and explanatory variables.

I agree with Sain and Furrer that “it would be interesting to see if the different types of coregionalization models are stationary and how they compare with each other in this respect.” While a stationary coregionalization model may be built by formulating stationary latent fields, an interesting question would be whether or how a stationary coregionalization model may be built from non-stationary latent fields. In addition, it would be interesting to know whether the “entanglement” of the spatial and non-spatial parameters in the coregionalization models, say the models of the Type II decomposition with full matrices  $\mathbf{C}$  and  $\mathbf{A}$  or their SVC counterparts, may give the MGMRFs the flexibility to model or approximate stationary or nearly stationary Gaussian fields.

The computational advantages of (stationary) GMRFs also motivated recent considerations of fitting (stationary) GMRFs to (stationary) Gaussian fields formulated through specifications of the covariance functions (Rue and Tjelmeland 2002; Cressie and Verzelen 2008; Lindgren et al. 2011). In this context, both the local and global properties of the GMRFs are important (Rue and Tjelmeland 2002). As noted in Rue and Tjelmeland (2002), one important question is whether a GMRF with a small neighborhood can approximate a Gaussian field with long correlation length. Figure 5 seems to indicate that the linear coregionalization MGMRF with element-wise HPs for an asymmetric matrix  $\mathbf{C}$  of spatial parameters, which control for conditional spatial dependencies and cross-dependencies in the latent MGMRF, may have the flexibility to approximate smooth multivariate Gaussian fields. A follow-up and more rigorous research into this perceived flexibility is necessary.

Rue and Tjelmeland (2002) indicate that local Markov random fields are able to fit global properties to some extent. Sain and Furrer mention the need of higher-order neighborhood structures for smoother fields. I agree with Sain and Furrer that extensions of the MGMRFs to higher-order neighborhood structures and associated Markovian dependence and independence may be conceptually straightforward but analytically and computationally complex. Nevertheless, formulation and implementation of coregionalization MGMRFs of higher-order neighborhood structures are

more manageable for  $p$ -variate GMRFs with  $p$  variable-specific spatial parameters or for separable MGMRFs with a general spatial parameter.

#### 4 Various approaches to model formulation and related applications

Sain and Furrer comment on the fact that, while the coregionalization framework unifies several lines of MGMRF development, “there is still no one unified model formulation that allows movement between the different approaches through some set of parameters.” They rightly correct me and show that the Sain et al. (2011) framework contains separable models. Indeed, if we free ourselves to allow the off-diagonal block matrix elements  $\beta_{ik}$  (when  $i \sim k$ ) in the Sain et al. MGMRF framework to be parameterized with both the spatial and non-spatial dependence parameters, the Sain et al. family of MGMRFs actually contains the MGMRFs of both Type I and II decompositions. To put it differently, the following *joint* precision matrix

$$\mathfrak{Q}_{\text{vec}(\zeta^\top)}^{\text{MGMRF}} = \mathbf{D}_m \otimes \boldsymbol{\Gamma} - (\mathbf{W}_U \otimes \mathbf{B} + \mathbf{W}_U^\top \otimes \mathbf{B}^\top) \quad (1)$$

(Equation (14) in the paper) represents a general formulation of the MGMRFs contained within the Sain et al. (2011), the Mardia (1988), and the linear coregionalization (MacNab 2016a, b) frameworks. Through various parameterizations of  $\mathbf{B}$  (eg.  $\mathbf{B} = \mathbf{B}(\mathbf{C}, \boldsymbol{\tau})$  or  $\mathbf{B} = \mathbf{B}(\mathbf{C}, \boldsymbol{\Gamma})$  or  $\mathbf{B} = \mathbf{B}(\mathbf{C}, \boldsymbol{\Sigma}^{1/2})$  or  $\mathbf{B} = \mathbf{B}(\mathbf{C}, \mathbf{A})$ ), specific MGMRFs of the Type I or II decomposition could be derived to have a precision matrix (1).

Martinez-Beneito comments on the need to better understand whether models produced from one approach can be reproduced from another approach. He also calls for better understanding of the different features of the models produced by the different approaches. The Sain and Furrer commentary and the above discussion offer some relevant new insights. For example, if we define a MGMRF by its joint precision matrix (1), the MGMRFs produced by the Mardia (1988) approach can be reproduced by the Sain et al. (2011) approach, and vice versa.

The models with a precision matrix (1) but with different parameterizations of  $\mathbf{B}$  are different MGMRFs with different partial correlation and cross-correlation matrix functions. They can also represent different conditionally formulated MGMRFs, one based on univariate conditionals and the other multivariate conditionals. For MGMRF estimation and inference, the different lines of model development and different model constructions also have had considerable influence on our choice for positive definiteness constraint and for hyperprior specification. As pointed out by Greco and Trivisano and discussed earlier, the observed differences between the various models, say, those presented in the present paper, may due in part to the different prior specifications for the model parameters. I agree with Martinez-Beneito on the appeal of casting the coregionalization MGMRFs within a matrix algebraic framework. For example, the spatially varying coregionalization MGMRFs presented in the paper can be seen as being built within a matrix algebraic framework. Indeed, the advantages of the Martinez-Beneito (2013) framework are well illustrated in Martinez-Beneito et al.

(2017), where the use of matrix theory and algebra for the formulations of complex M-models, and the associated statistical computations, is presented.

In general, the challenges in constructing, constraining, and estimating a MGMRF differ considerably depending on whether we pursue a separable or non-separable model. If a non-separable model is considered, then a model with a diagonal matrix of spatial parameters is, in general, more readily constrained and estimated, compared to its counterparts with a full matrix of spatial parameters. My own experiences, and the results presented in recent literature, also correspond with Martínez-Beneito's comment that, at least in the context of multivariate disease mapping, MGMRFs with a full matrix of spatial parameters may not be necessary or may be over-parameterized, particularly for data of rare events.

MGMRFs with a full matrix of spatial dependence parameters may be useful when the goal is estimation and inference on multivariate spatial dependencies. For example, in the Sain et al. (2011) study, the motivating example for their bivariate MGMRF proposal was to model and draw inference on asymmetric local dependencies between two climate variables: temperature and precipitation. The pair-wise conditional asymmetric spatial dependencies are quantified in relation to the variables and to the site labeling. In some applications, this may be an appealing feature of the MGMRFs. For example, complex and diverse interaction structures may be modeled by varying the neighborhood structures, the labeling of the neighbor sets, and the parameters in the MGMRFs. In the contexts of image analysis and restoration, computer vision, social network analysis, and spatial data fusion, these MGMRFs may be potentially useful for modeling and learning complex and varied local patterns and features of dependencies and interactions.

Indeed, there is a lot to learn about the various MGMRF constructions. A good understanding of the various approaches to formulating MGMRGs should enable us to develop subject-matter-specific models that provide principled ways to express dependency and interaction structures. I agree with Sain and Furrer that developing objective procedures and practical guidance for choosing between competing models is an area of ongoing and necessary research and progress. Potential utilities of the various MGMRF constructions may be better explored as we succeed in tackling the computational challenges in statistical estimation and inference. Overcoming these challenges may also open new frontiers for MGMRF development and application.

## 5 Statistical computation

As mentioned in the paper, the currently available computational methods and tools for Bayesian hierarchical MGMRF models primarily use Gibbs or Metropolis-within-Gibbs sampling algorithms that capitalize on the *conditional probability formulations* of (latent) Markov random fields (Besag et al. 1991; Besag and Green 1993). The full conditionals facilitate relatively simple programming for location-wise or variable- and location-wise posterior sampling, often requiring little or no matrix algebra. The main disadvantage of these component-wise Gibbs sampling methods is that the MCMC simulations can be impractically slow and the computational costs may be prohibitive for datasets with a large number of sites (i.e., areal units) and/or a large number of vari-

ables. Nevertheless, these computational tools are useful for modestly sized datasets and have enabled us to gain deeper knowledge about the conditionally formulated models discussed in the present paper.

While it contains limited mathematics tools, the WinBUGS (or OpenBUGS) free-ware offers a user friendly and accessible interface for Bayesian analysis of the majority of the MGMRFs available to date. As illustrated in the recent literature and in the present paper, WinBUGS may still be quite useful to statisticians and practitioners who wish to use, learn, and test these MGMRFs in real-life applications, at least in the near future.

I agree with Greco–Trivisano and Sain–Furrer that writing computer code and packages outside WinBUGS, say, for a “ready-to-use Bayesian software environment,” would be a worthwhile effort and can be essential for computational flexibility, efficiency, and scalability. In the pursuit of this effort, alternative computational methods and tools may be developed by tapping into sparse matrix methods that are available in software of high-level programming language, such as the R (<https://www.r-project.org>), Python (<https://www.python.org>), and MATLAB (<https://www.mathworks.com/products/matlab.html>). For example, an R-package may be developed for existing computational methods and traditional component-wise or block Gibbs samplers (Rue 2001; Rue and Held 2005). New Gibbs updating strategies for computationally efficient posterior sampling on large lattices (Brown et al. 2017; Marcotte and Allard 2018) may also be explored by programming in R, Python, or MATLAB.

There are also several less-explored computational options that can take the advantage of sparse MGMRF precision matrix. For example, instead of using the Gibbs sampler for fully Bayesian hierarchical inference involving MGMRF, we may explore the possibility of developing an R-package for the so-called hybrid Monte Carlo algorithm, also known as the HMC or the Hamiltonian MC algorithm (Neal 1996; MacNab 2003a, b; MacNab et al. 2004; Gustafson et al. 2004; Girolami and Calderhead 2011). If successful, the R-package may provide a tool for MCMC sampling of complex multivariate posteriors, say, for the generalized linear mixed (GLMM) models with the SVC priors discussed in the paper. In the context of Bayesian disease mapping and ecological regression, my earlier works in this direction (MacNab 2003a, b; MacNab et al. 2004; Gustafson et al. 2004) explored GMRF estimation for mostly large datasets (MacNab 2003a, b). Compared to the component-wise Gibbs sampler, an adequately tuned HMC algorithm may facilitate computationally more efficient joint posterior sampling of correlated (latent) components, such as correlated random effects in GLMMs. For MGMRFs, a computational challenge is again the tuning of user-specified parameters that (i) control the step size for proposal distribution and (ii) determine a desired number of Monte Carlo runs. Recent works considered optimal tuning (Beskos et al. 2013) or automatic tuning of the HMC parameters (Hoffman and Gelman 2014). These lines of research are important and should make the HMC algorithm more accessible.

We may also tap into the existing tools for Bayesian or approximate Bayesian estimation and inference. For example, the Hamiltonian Monte Carlo sampling tools offered by Stan interfaces (Stan Development Team 2016), such as **rstan** for R, **PyStan** for Python, and **MatlabStan** for MATLAB, may be explored and utilized. The R-package for stochastic gradient MCMC, **sgmcmc**, may also be considered or expanded

as a computational option for large datasets. Another option is to access and improve the well-known *Integrated Nested Laplace Approximations* (INLA) tool in R, the **R-INLA**, for approximate Bayesian inference, perhaps for MGMRFs of small or modest  $p$  and a modest number of hyperparameters; see Rue et al. (2017) for a recent review on approximate Bayesian computing with INLA.

Sain and Furrer comment on likelihood estimation as a means to explore and address issues concerning (i) choice of parameterization and (ii) potential impact of transformations or constraints for parameters on estimation. These and similar issues may also be explored and addressed within a Bayesian hierarchical inferential framework using efficient Bayesian tools. Nevertheless, likelihood-based estimation methods, such as the pseudolikelihood approach (Besag 1974, 1975), (penalized) maximum likelihood methods (Dempster 1977; Fessler and Hero 1995; Descombes et al. 1999; Zammit-Mangion and Rougier 2018), penalized quasi-likelihood methods (Breslow and Clayton 1993; Guha et al. 2009; Huque et al. 2018), or suitable variations, may indeed be useful options. Likelihood approaches to hierarchical MGMRF estimation typically involve (i) manipulations of sparse MGMRF precision matrices, (ii) iterative procedures, and (iii) careful and adequate quantification of estimation uncertainty (Ainsworth and Dean 2006; MacNab et al. 2004; MacNab and Lin 2009; Guha et al. 2009).

Variational inference (de Freitas et al. 2001; Kucukelbir et al. 2015, 2017; Blei et al. 2017 (a recent review); Zhang et al. 2018), composite likelihood methods (see Varin et al. 2011; Larribe and Fearnhead 2011 for recent reviews), and parallel computing (Gonzalez et al. 2011; Brown et al. 2017; Castruccio and Genton 2018) are also potential options to be explored and utilized for analyzing data on large lattices.

## References

- Ainsworth LM, Dean CB (2006) Approximate inference for disease mapping. *Comput Stat Data Anal* 50:2552–2570
- Besag J (1972) On the correlation structure of some two-dimensional stationary processes. *Biometrika* 59:43–48
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J R Stat Soc Ser B* 36: 192–236
- Besag J (1975) Statistical analysis of non-lattice data. *J R Stat Soc Ser D (Stat)* 24(3):179–195
- Besag J, Green PJ (1993) Spatial statistics and Bayesian computation. *J R Stat Soc Ser B (Methodol)* 55(1):25–37
- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43:1–21
- Beskos A, Pillai N, Roberts G, Sanz-Serna J, Stuart A (2013) Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* 19(5A):1501–1534
- Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: a review for statisticians. *J Am Stat Assoc* 112(518):859–877
- Botella-Rocamora P, Martínez-Beneito MA, Banerjee S (2015) A unifying modelling framework for highly multivariate disease mapping. *Stat Med* 34(9):1548–1559
- Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88:925
- Brown AD, McMahan CS, Watson SC (2017) Sampling strategies for fast updating of Gaussian Markov random fields. Preprint [arXiv:1702.05518](https://arxiv.org/abs/1702.05518)

- Castruccio S, Genton M (2018) Principles for statistical inference on big spatio-temporal data from climate models. *Stat Probab Lett* 136:92–96
- Cressie N, Verzelen N (2008) Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. *Comput Stat Data Anal* 52(5):2794–2807
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39(1):1–38
- de Freitas N, Hojen-Sorensen P, Jordan MI, Russell S (2001) Variational MCMC. In: Proceedings of the seventeenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., San Francisco, pp 120–127
- Descombes X, Sigelle M, Préteux F (1999) Estimating Gaussian Markov random field parameters in a nonstationary framework: application to remote sensing imaging. *IEEE Trans Image Process* 8(4):490–503
- Fessler JA, Hero AO (1995) Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Trans Image Process* 4(10):1417–1429
- Girolami M, Calderhead B (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussions). *J R Stat Soc Ser B (Methodol)* 73:123214
- Gonzalez JE, Low Y, Gretton A, Guestrin C (2011) Parallel Gibbs sampling: from colored fields to thin junction trees. In: Journal of machine learning research: proceedings of the 14th international conference on artificial intelligence and statistics (AISTATS), pp 324–332
- Guha S, Ryan L, Morara M (2009) Gauss–seidel estimation of generalized linear mixed models with application to poisson modeling of spatially varying disease rates. *J Comput Graph Stat* 18(4):818–837
- Gustafson P, MacNab YC, Wen S (2004) The value of derivatives and random walk suppression in Markov chain Monte Carlo algorithms. *Stat Comput* 14(1):23–38
- Huque MH, Anderson C, Walton R, Woolford S, Ryan L (2018) Smooth individual level covariates adjustment in disease mapping. *Biomet J* 60:597–615
- Hoffman MD, Gelman A (2014) The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15(1):15931623
- Kucukelbir A, Ranganath R, Gelman A, Blei D (2015) Automatic variational inference in Stan. In: Neural information processing systems, pp 568–576
- Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei D (2017) Automatic differentiation variational inference. *J Mach Learn Res* 18:1–45
- Larribe F, Fearnhead P (2011) On composite likelihoods in statistical genetics. *Stat Sin* 21(1):43–69
- Lindgren F, Rue H, Lindstrom J (2011) An explicit link between Gaussian fields and Gaussian Markov random field: The stochastic partial differential equation approach. *J R Stat Soc Ser B* 73:423–498
- MacNab YC (2003a) Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics* 59:305–316
- MacNab YC (2003b) Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease. *Stat Med* 22(10):1761–73
- MacNab YC (2016a) Linear models of coregionalization for multivariate lattice data: a general framework for coregionalized multivariate CAR models. *Stat Med* 35:3827–3850
- MacNab YC (2016b) Linear Models of coregionalization for multivariate lattice data: Order-dependent and order-free MCARs. *Stat Methods Med Res* 25(4):1118–1144
- MacNab YC, Farrell PJ, Gustafson P, Wen S (2004) Estimation in Bayesian disease mapping. *Biometrics* 60:865–873
- MacNab YC, Lin Y (2009) On empirical Bayes penalized quasilielihood inference in GLMMs and in Bayesian disease mapping and ecological modeling. *Comput Stat Data Anal* 53(8):2950–2967
- MacNab YC (2018). Positive definiteness constraints and shrinkage estimation for multivariate Gaussian Markov random fields. (Unpublished manuscript)
- Marcotte D, Allard D (2018) Gibbs sampling on large lattice with GMRF. *Comput Geosci* 111:190–199
- Mardia KV (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J Multivar Anal* 24:265–284
- Martinez-Beneito MA (2013) A general modelling framework for multivariate disease mapping. *Biometrika* 100(3):539–553
- Martinez-Beneito MA, Botella-Rocamora P, Banerjee S (2017) Towards a multi-dimensional Approach to Bayesian Disease Mapping. *Bayesian Anal* 12(1):239–259
- Neal RM (1996) Bayesian learning for neural networks. Springer, New York

- Rue H (2001) Fast sampling of Gaussian Markov random fields. *J R Stat Soc Ser B* 63:325–338
- Rue H, Tjelmeland H (2002) Fitting Gaussian Markov random fields to Gaussian fields. *Scand J Stat* 29:31–49
- Rue H, Held L (2005) *Gaussian Markov random fields: theory and applications*. Chapman & Hall, New York
- Rue H, Riebler A, Srbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian computing with INLA: a review. *Ann Rev Stat Appl* 4:395–421
- Sain SR, Furrer R, Cressie N (2011) A spatial analysis of multivariate output from regional climate models. *Ann Appl Stat* 5(1):150–175
- Stan Development Team (2016) *Stan modeling language users guide and reference manual*, Versuub 2.14.0. <http://mc-stan.org/>. Accessed May 2018
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Stat Sin* 21(1):5–42
- Zammit-Mangion A, Rougier J (2018) A sparse linear algebra algorithm for fast computation of prediction variances with Gaussian Markov random fields. *Comput Stat Data Anal* 123:116–130
- Zhang C, Shahbaba B, Zhao H (2018) Variational Hamiltonian Monte Carlo via score matching. *Bayesian Anal* 13(2):485–506