Stoch Environ Res Risk Assess (2018) 32:749–770
https://doi.org/10.1007/s00477-017-1405-0

ORIGINAL PAPER

In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results

T. Goicoa[1,2,4] · A. Adin[1,2] · M. D. Ugarte[1,2] · J. S. Hodges[3]

Published online: 30 March 2017

Abstract Disease mapping studies the distribution of relative risks or rates in space and time, and typically relies on generalized linear mixed models (GLMMs) including fixed effects and spatial, temporal, and spatio-temporal random effects. These GLMMs are typically not identifiable and constraints are required to achieve sensible results. However, automatic specification of constraints can sometimes lead to misleading results. In particular, the penalized quasi-likelihood fitting technique automatically centers the random effects even when this is not necessary. In the Bayesian approach, the recently-introduced integrated nested Laplace approximations computing technique can also produce wrong results if constraints are not well-specified. In this paper the spatial, temporal, and spatio-temporal interaction random effects are reparameterized using the spectral decompositions of their precision matrices to establish the appropriate identifiability constraints. Breast cancer mortality data from Spain is used to illustrate the ideas.

Keywords Breast cancer · INLA · Leroux CAR prior · PQL · Space-time interactions

✉ M. D. Ugarte
lola@unavarra.es

1 Department of Statistics and O.R., Public University of Navarre, Pamplona, Spain

2 Institute for Advanced Materials (InaMat), Public University of Navarre, Pamplona, Spain

3 Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, USA

4 Research Network on Health Services in Chronic Diseases (REDISSEC), Madrid, Spain

1 Introduction

Statistical techniques for disease mapping have developed tremendously in the last few years. The availability of information from modern registers with high quality data recorded for many years and regions has brought about new goals and challenges, which in turn have made possible the development of new and more flexible statistical models, faster and less computationally demanding fitting techniques, and new free software to implement these advances. These methodological developments are now ready for policy makers, epidemiologists, health researchers, and health professionals to use in a more or less automated form. However, this abundance of ready-to-use statistical resources can lead to errors and misleading results when analyzing mortality or incidence data in space and time due, among other causes, to incorrect specification of identifiability constraints, which standard software usually fixes at default values.

Research into spatial and spatio-temporal disease mapping has been carried out within a Bayesian framework, with generalized linear mixed models (GLMM) playing a major role. Two main approaches have been followed for model fitting and inference, the empirical Bayes (EB) and fully Bayes (FB) approaches. Model fitting and estimation in the EB approach commonly rely on penalized quasi-likelihood (PQL) (see Breslow and Clayton 1993), while the FB approach usually uses Markov chain Monte Carlo (McMC) techniques (Gilks 2005). In addition to McMC, a new strategy based on integrated nested Laplace approximations (INLA) has recently been derived for estimating posterior quantities of interest (Rue et al. 2009). This technique is becoming very popular in disease mapping because it allows fairly complex space-time models to be fit much more quickly than McMC.

The GLMMs traditionally used in disease mapping are not identifiable (Gelfand and Sahu 1999) and although some identifiability problems have been dealt with, this matter deserves further attention and needs clarification for practitioners. For example, one of the first identifiability concerns arose with the work by Besag et al. (1991). They proposed an areal model (the BYM model) for the log-relative risks of a disease considering two random area effects: one with an exchangeable distribution and one with an intrinsic conditional autoregressive (ICAR) distribution. The ICAR distribution is specified conditionally and the parameters are uniquely determined up to an additive constant, so the overall intercept is implicit in the ICAR specification. Hence, if the model includes an explicit intercept as well, the model is not identified. The solution is to omit the explicit intercept or to add sum-to-zero constraints for the random effects.

Counts in space and time demand more flexible models to unveil the underlying geographical patterns and their temporal evolution. However, as terms are added to the model, identifiability problems arise. The literature in spatio-temporal disease mapping is abundant, describing different models with parametric (see for example Bernardinelli et al. 1995; Ugarte et al. 2009) as well as non parametric trends (Knorr-Held and Besag 1998). A key research paper is Knorr-Held (2000), which specifies spatio-temporal models including four different types of spatio-temporal interactions. In these models, identifiability problems arise because the overall level can be absorbed by either the spatial or the time main effect, and the interaction terms are confounded with the main effects. A different type of spatio-temporal model in disease mapping combines CAR spatial random effects with temporal trends based on B-splines (see, e.g., MacNab and Gustafson 2007; MacNab 2007). In most of this research, sum-to-zero constraints are considered as guaranteeing model identifiability but no clear guidance is given about why these constraints have to be considered. The foregoing papers took a FB approach using McMC, while other papers took the EB approach, using PQL for model fitting. For example, Ugarte et al. (2010) consider spatio-temporal CAR models and P-spline models from an EB perspective to study brain cancer mortality in Spain, and Etxeberria et al. (2014) consider spatio-temporal CAR and P-spline models for smoothing and forecasting mortality risks. However, identifiability issues have not received much attention because the PQL method automatically centers the spatial, temporal, and spatio-temporal random effects, that is, automatically imposes sum-to-zero constraints.

This paper considers space-time disease mapping models including an overall risk level (intercept) and spatial, temporal, and spatio-temporal random effects. In particular, conditional autoregressive (CAR) spatial random effects, first- or second-order random walks for time, and the corresponding space-time interactions are considered. To deal with identifiability problems, models will be reparameterized using the spectral decomposition of the precision matrices to remove the combinations of the random effects that are in the span of the fixed effects as in Reich et al. (2006) and Hodges and Reich (2010). These authors treat the problem as a collinearity issue and restrict the spatial random effects to the subspace orthogonal to the fixed effects. Using this reparameterization it is easy to see that deleting repeated columns in certain model matrices is equivalent to specific sum-to-zero constraints. This is important for practitioners as some statistical software requires the specification of such constraints. In particular, we will focus on specifying constraints in the `R-INLA` package for approximate Bayesian inference, which has become popular because it provides quick fits of complex models.

The rest of the paper is laid out as follows. Section 2 reviews a simple spatial model and a more general spatio-temporal model with identifiability problems. Section 3 considers a reparameterization to make the models identifiable. Section 4 provides insight into model estimation and the use of linear constraints with PQL. Section 5 illustrates the previous sections' results using a case study. The paper closes with a discussion.

2 Spatial and spatio-temporal models in disease mapping

This section briefly reviews spatial and spatio-temporal disease mapping models to highlight the identifiability problems arising in this field.

Consider a large domain (let us say a country) divided into small areas (for example provinces or counties) that will be labelled by $i = 1, \ldots, S$, and denote by Y_i the number of deaths (or incident cases) in the ith small area. Then conditional on the relative risk r_i, Y_i is assumed to be Poisson distributed with mean $\mu_i = e_i r_i$, where e_i is the number of expected cases. That is

$$Y_i | r_i \sim Pois(\mu_i = e_i r_i), \;\; \log \mu_i = \log e_i + \log r_i.$$

Here $\log e_i$ is an offset and $\log r_i$ is modeled as

$$\log r_i = \eta + \xi_i, \tag{1}$$

where η is an overall risk and ξ_i is the spatial random effect. An intrinsic conditional autoregressive (ICAR) prior is considered for the vector of spatial effects $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_S)'$. Namely,

$$\boldsymbol{\xi} \sim N(\mathbf{0}, \sigma^2_{\xi} \mathbf{Q}^{-}_{\xi}),$$

where $^-$ indicates the Moore-Penrose inverse of a matrix, and $\mathbf{Q}_\xi$ is the $S \times S$ spatial neighbourhood matrix with (i, j) element defined as $\mathbf{Q}_{\xi(ij)} = -1$ if areas i and j are neighbours and 0 otherwise. The ith diagonal element equals the number of neighbours of the ith region. Typically, two regions are neighbours if they share a common border. Clearly, $\sum_j \mathbf{Q}_{\xi(ij)} = 0, \forall i$, that is $\mathbf{Q}_\xi \mathbf{1}_S = \mathbf{0}$, where $\mathbf{1}_S$ is a vector of ones of length S, and the intercept is implicit in the ICAR specification as will be shown in Sect. 3.1. Hence, an identifiability problem arises for the intercept. The problem can be solved either by deleting the intercept or by imposing sum-to-zero constraints $\sum_i \xi_i = 0$ (see for example, Eberly and Carlin 2000).

Other priors for the spatial random effects have been proposed. Leroux et al. (1999) considered the following specification (LCAR hereafter in the paper) taking account of spatially structured and unstructured variability

$$\boldsymbol{\xi} \sim N(\mathbf{0}, \sigma_\xi^2 \mathbf{D}^{-1}),\ \mathbf{D} = (\lambda_\xi \mathbf{Q}_\xi + (1 - \lambda_\xi)\mathbf{I}_\xi), \tag{2}$$

where $\lambda_\xi \in [0, 1]$ is a spatial smoothing parameter and $\mathbf{I}_\xi$ is an $S \times S$ identity matrix. If $\lambda_\xi = 1$, Model (2) becomes the ICAR distribution. If $0 \leq \lambda_\xi < 1$, the matrix $\mathbf{D}$ is of full rank, but a mild identifiability issue (some might prefer "confounding issue") still remains, as we will see in Sect. 3.

Suppose now that for each small area i, data are available for different time periods labelled by $t = 1, \ldots, T$. Then, conditional on the relative risk r_{it}, the count of events in region i at time t, Y_{it}, is assumed to follow a Poisson distribution with mean $\mu_{it} = e_{it} r_{it}$, where e_{it} is the number of expected events. That is

$$Y_{it} | r_{it} \sim Pois(\mu_{it} = e_{it} r_{it}),\ \log \mu_{it} = \log e_{it} + \log r_{it}.$$

The term $\log r_{it}$ can include spatial and temporal random effects additively, as well as space-time interactions. Let us focus first on the following spatio-temporal additive model

$$\log r_{it} = \eta + \xi_i + \gamma_t, \tag{3}$$

where here the vector of temporal random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)'$ is assumed to follow a random walk of first (RW1) or second (RW2) order, that is,

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma_\gamma^2 \mathbf{Q}_\gamma^-),$$

where the $T \times T$ matrix $\mathbf{Q}_\gamma$ has rank deficiency equal to 1 or 2 for a RW1 and a RW2 respectively (see Rue and Held 2005, chap. 3). If the temporal random effect is assumed to follow a RW1 distribution, then $\mathbf{Q}_\gamma \mathbf{1}_T = \mathbf{0}$, where $\mathbf{1}_T$ is a vector of ones of length T, and the intercept is implicit in the RW1, leading to an identifiability problem. As in the spatial case, the problem can be solved either by deleting the intercept or by imposing the sum-to-zero constraint $\sum_t \gamma_t = 0$. If the spatial random effect follows an ICAR distribution, we can delete the intercept and impose sum-to-zero constraints on the spatial or the temporal random effects, or leave the intercept and impose sum-to-zero constraints on both the spatial and the temporal random effects. If the temporal random effect is distributed as RW2, then $\mathbf{Q}_\gamma \mathbf{1}_T = \mathbf{0}$ and $\mathbf{Q}_\gamma \mathbf{t}^* = \mathbf{0}$, where $\mathbf{t}^* = (1, 2, \ldots, T)'$, so the slope in time is implicit in the RW2 specification. In contrast to the intercept, the model does not include a linear trend, so no extra constraints are needed.

Spatio-temporal models including area and time effects additively may be very restrictive in practice, so interaction terms are usually added to Model (3). Knorr-Held (2000) proposes four types of interactions, namely Type I (interaction random effects with any spatial or temporal structure), Type II (interaction random effects spatially unstructured but temporally correlated), Type III (interaction random effects spatially correlated but with no temporal structure), and finally Type IV (interaction random effects spatially and temporally correlated). Here we focus on Type IV interactions, the most complex type (see "Appendix 1" for full details about constraints for Type IV interactions, and "Appendix 2" for Type I, Type II, and Type III interactions). Then Model (3) becomes

$$\log r_{it} = \eta + \xi_i + \gamma_t + \delta_{it}, \tag{4}$$

where the vector of spatio-temporal random effects $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{S1}, \ldots, \delta_{1T}, \ldots, \delta_{ST})'$ is assumed to follow the multivariate normal distribution

$$\boldsymbol{\delta} \sim N\big(\mathbf{0}, \sigma_\delta^2 (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^-\big).$$

If $\mathbf{Q}_\xi$ specifies an ICAR model, then the rank of the matrix $\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi$ is $(T-1) \times (S-1)$ or $(T-2) \times (S-1)$ if $\boldsymbol{\gamma}$ follows a RW1 or a RW2 respectively. Consequently, the rank deficiency is $T + S - 1$ (RW1) or $T + 2S - 2$ (RW2). Note that $(\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)\mathbf{1}_{TS} = \mathbf{0}$, where $\mathbf{1}_{TS}$ is a vector of ones of length TS, leading to an identifiability problem with the intercept. In addition, the interaction term is confounded with the main effects, creating further identifiability issues. In the next section, we reparameterize the models using the spectral decomposition of the precision matrices of the random effects to solve these identifiability problems.

3 Model reparameterization

In this section, the random effects are transformed using appropriate matrices to express them as independent Gaussian random effects. Deleting repeated columns in the design matrices circumvents the identifiability issues, which implies suitable constraints.

3.1 Spatial model

Consider the spatial Model (1) in matrix form

$$\log \mathbf{r} = \mathbf{1}_S \eta + \xi, \quad \xi \sim N\left(\mathbf{0}, \sigma_\xi^2 \mathbf{Q}_\xi^-\right), \tag{5}$$

where $\mathbf{r} = (r_1, \ldots, r_S)'$. The neighbourhood matrix $\mathbf{Q}_\xi$ has rank deficiency 1 assuming the spatial domain is connected. Consider the spectral decomposition of $\mathbf{Q}_\xi$,

$$\mathbf{Q}_\xi = \mathbf{U}_\xi \boldsymbol{\Sigma}_\xi \mathbf{U}'_\xi = [\mathbf{U}_{\xi n} : \mathbf{U}_{\xi r}] \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Sigma}}_\xi \end{pmatrix} \begin{bmatrix} \mathbf{U}'_{\xi n} \\ \mathbf{U}'_{\xi r} \end{bmatrix}, \tag{6}$$

where $\mathbf{U}_\xi = [\mathbf{U}_{\xi n} : \mathbf{U}_{\xi r}]$ is an orthogonal matrix with columns the eigenvectors of $\mathbf{Q}_\xi$, $\mathbf{U}_{\xi n} = \mathbf{1}_S$ (up to a normalizing constant) and $\mathbf{U}_{\xi r}$ are the matrices of eigenvectors having null and non-null eigenvalues respectively, and $\widetilde{\boldsymbol{\Sigma}}_\xi$ is a diagonal matrix with the non-null eigenvalues of $\mathbf{Q}_\xi$ in the main diagonal. Then, as $\mathbf{U}_\xi$ is orthogonal,

$$\boldsymbol{\xi} = \mathbf{U}_\xi \mathbf{U}'_\xi \boldsymbol{\xi} = [\mathbf{U}_{\xi n} : \mathbf{U}_{\xi r}] \begin{bmatrix} \mathbf{U}'_{\xi n} \\ \mathbf{U}'_{\xi r} \end{bmatrix} \boldsymbol{\xi}.$$

If we define

$$\mathbf{X} = \mathbf{U}_{\xi n} = \mathbf{1}_S, \beta_\xi = \mathbf{U}'_{\xi n} \boldsymbol{\xi} = \mathbf{1}'_S \boldsymbol{\xi}$$
$$\mathbf{Z} = \mathbf{U}_{\xi r}, \quad \boldsymbol{\alpha}_\xi = \mathbf{U}'_{\xi r} \boldsymbol{\xi},$$

then

$$\boldsymbol{\xi} = \mathbf{X} \beta_\xi + \mathbf{Z} \boldsymbol{\alpha}_\xi,$$

where β_ξ plays the role of a fixed effect and the spatial Model (5) can be reformulated as

$$\log \mathbf{r} = \mathbf{1}_S \eta + \mathbf{1}_S \beta_\xi + \mathbf{U}_{\xi r} \boldsymbol{\alpha}_\xi, \quad \boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma_\xi^2 \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right). \tag{7}$$

Note that the reparameterized Model (7) has two intercepts, revealing the identifiability problem. Consequently, removing or setting to zero the intercept β_ξ makes the model identifiable, and now the precision matrix of the reparameterized random effect has full rank. Setting β_ξ to zero leads to the usual sum-to-zero constraint $\sum_{i=1}^S \xi_i = 0$, as $\beta_\xi = \mathbf{1}'_S \boldsymbol{\xi} = \sum_{i=1}^S \xi_i$. The identifiable spatial model is then

$$\log \mathbf{r} = \mathbf{1}_S \eta + \mathbf{U}_{\xi r} \boldsymbol{\alpha}_\xi, \quad \boldsymbol{\alpha} \sim N\left(\mathbf{0}, \sigma_\xi^2 \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right).$$

If the prior for the spatial random effect is the LCAR given in Eq. (2), the covariance matrix is of full rank whenever $0 \le \lambda_\xi < 1$, but a confounding problem—a milder version of the ICAR's identifiability problem—still remains. In particular, the matrix $\mathbf{D}$ has spectral decomposition

$$\mathbf{D} = (\lambda_\xi \mathbf{Q}_\xi + (1 - \lambda_\xi)\mathbf{I}_\xi) = \left(\lambda_\xi \mathbf{U}_\xi \boldsymbol{\Sigma}_\xi \mathbf{U}'_\xi + (1 - \lambda_\xi)\mathbf{I}_\xi\right)$$
$$= \mathbf{U}_\xi (\lambda_\xi \boldsymbol{\Sigma}_\xi + (1 - \lambda_\xi)\mathbf{I}_\xi)\mathbf{U}'_\xi,$$

because $\mathbf{U}_\xi$ is orthogonal, so that $\mathbf{D}$ and $\mathbf{Q}_\xi$ have the same eigenvectors but different eigenvalues. Defining $\mathbf{X} = \mathbf{U}_{\xi n} = \mathbf{1}_S$, $\beta_\xi = \mathbf{U}'_{\xi n} \xi = \mathbf{1}'_S \xi$, $\mathbf{Z} = \mathbf{U}_{\xi r}$, and $\boldsymbol{\alpha} = \mathbf{U}'_{\xi r} \boldsymbol{\xi}$ as before, the spatial random effect can be expressed as $\boldsymbol{\xi} = \mathbf{X} \beta_\xi + \mathbf{Z} \boldsymbol{\alpha}$ and again we have a redundant intercept. As a reviewer suggested, the identifiability issues in the ICAR and LCAR models are not identical, and we now clarify the difference. For $\lambda \in [0, 1)$, all eigenvalues of $\mathbf{D}$ are positive and $\beta_\xi \sim N\left(0, \sigma_\xi^2 d_{11}\right)$ is partially identified by its prior, where d_{11} is the eigenvalue of $\mathbf{D}$ corresponding to the eigenvector $\mathbf{1}_S$. However, it is substantively vacuous to distinguish the two intercepts η and β_ξ. The only thing that matters substantively is their sum, and the allocation of that sum between the fixed and random effect is arbitrary. If the LCAR model is considered for the random effects, in a Bayesian analysis the point estimates of the two intercepts can be made to take any values adding to the aforementioned sum, by selecting the right prior for the fixed effect. In a non-Bayesian analysis, where the fixed-effect intercept implicitly has an improper flat prior and the LCAR's intercept has positive prior precision, the LCAR's intercept will always be estimated to be zero (this is shown in Sect. 4). The implication is that keeping the LCAR's intercept confounds the fixed-effect intercept and inflates its variance (see "Appendix 3") even though it is identified. Thus there is never an advantage to retaining the separate intercept arising from the LCAR, while omitting it averts variance inflation. Therefore, it should be omitted. That is why, in practical terms, we treat both the ICAR and LCAR cases similarly.

3.2 Spatio-temporal model

We now consider the spatio-temporal Model (4), with the LCAR prior (2) for the spatial random effect; we consider this prior as it takes account of both structured and unstructured variability. The matrix form of this model is

$$\log \mathbf{r} = \mathbf{1}_{TS} \eta + (\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi} + (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} + \mathbf{I}_\delta \boldsymbol{\delta}, \tag{8}$$

where $\mathbf{r} = (r_{11}, \ldots, r_{S1}, \ldots, r_{1T}, \ldots, r_{ST})'$, and $\mathbf{I}_\gamma$ and $\mathbf{I}_\delta$ are $T \times T$ and $TS \times TS$ identity matrices respectively. The temporal main effect $\boldsymbol{\gamma}$ is assumed to follow a RW1 or a RW2, and the interaction random effect is assumed to be completely structured in space and time, that is, $\boldsymbol{\delta} \sim N\left(\mathbf{0}, \sigma_\delta^2 (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^-\right)$. Now consider the spectral decomposition of $\mathbf{Q}_\xi$ given by (6), and the spectral decomposition of $\mathbf{Q}_\gamma$

$$\mathbf{Q}_\gamma = \mathbf{U}_\gamma \boldsymbol{\Sigma}_\gamma \mathbf{U}'_\gamma = \left[\mathbf{U}_{\gamma n} : \mathbf{U}_{\gamma r}\right] \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Sigma}}_\gamma \end{pmatrix} \begin{bmatrix} \mathbf{U}'_{\gamma n} \\ \mathbf{U}'_{\gamma r} \end{bmatrix},$$

where $\mathbf{U}_\gamma = \left[\mathbf{U}_{\gamma n} : \mathbf{U}_{\gamma r}\right]$ is the $T \times T$ matrix of eigenvectors, $\mathbf{U}_{\gamma n}$ is the $T \times d$ matrix of eigenvectors having null eigenvalue, $\mathbf{U}_{\gamma r}$ is the $T \times (T-d)$ matrix of eigenvectors having non-null eigenvalues, and $\widetilde{\boldsymbol{\Sigma}}_\gamma$ is a diagonal matrix with the non-null eigenvalues in the main diagonal. Here d is the rank deficiency of matrix $\mathbf{Q}_\gamma$. If the distribution of the temporal random effect is a RW1, then the rank deficiency of $\mathbf{Q}_\gamma$ is $d = 1$, $\mathbf{U}_{\gamma n} = \mathbf{1}_T$ (up to a normalizing constant), and $\mathbf{U}_{\gamma r}$ is a $T \times (T-1)$ matrix. If the temporal random effect is distributed according to a RW2, the rank deficiency of $\mathbf{Q}_\gamma$ is $d = 2$, and $\mathbf{U}_{\gamma n} = [\mathbf{1}_T : \mathbf{t}^*]$, where $\mathbf{t}^* = (1, 2, \ldots, T)'$ up to a normalizing constant, and $\mathbf{U}_{\gamma r}$ is a $T \times (T-2)$ matrix. The spectral decomposition of $\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi$ can be expressed as

$$\mathbf{Q}_\delta = \mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi = \mathbf{U}_\delta \boldsymbol{\Sigma}_\delta \mathbf{U}'_\delta = [\mathbf{U}_{\delta n} : \mathbf{U}_{\delta r}] \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Sigma}}_\delta \end{pmatrix} \begin{bmatrix} \mathbf{U}'_{\delta n} \\ \mathbf{U}'_{\delta r} \end{bmatrix},$$

where $\mathbf{U}_\delta = [\mathbf{U}_{\delta n} : \mathbf{U}_{\delta r}]$ is the matrix of eigenvectors, $\mathbf{U}_{\delta n}$ is the matrix of eigenvectors having null eigenvalue, $\mathbf{U}_{\delta r}$ is the matrix of eigenvectors having non-null eigenvalues, and $\widetilde{\boldsymbol{\Sigma}}_\delta = \widetilde{\boldsymbol{\Sigma}}_\gamma \otimes \widetilde{\boldsymbol{\Sigma}}_\xi$ is a diagonal matrix with the non-null eigenvalues in the main diagonal. The matrix with eigenvectors spanning the null space can be expressed in terms of the eigenvectors spanning the null spaces of $\mathbf{Q}_\gamma$ and $\mathbf{Q}_\xi$, that is

$$\mathbf{U}_{\delta n} = [\mathbf{U}_{\gamma n} \otimes \mathbf{U}_{\xi n} : \mathbf{U}_{\gamma n} \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi n}].$$

Similarly,

$$\mathbf{U}_{\delta r} = [\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r}].$$

The key now is to define transformations so that the spatio-temporal Model (8) is reformulated to achieve identifiability. Define the transformation matrices as $\mathbf{U}_\gamma$ and $\mathbf{U}_\delta$ such that

$$(\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_\gamma \mathbf{U}'_\gamma \boldsymbol{\gamma} = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\left[\mathbf{U}_{\gamma n} : \mathbf{U}_{\gamma r}\right] \begin{bmatrix} \mathbf{U}'_{\gamma n} \\ \mathbf{U}'_{\gamma r} \end{bmatrix} \boldsymbol{\gamma}$$
$$= \left[\mathbf{X}_\gamma : \mathbf{Z}_\gamma\right] \begin{bmatrix} \beta_\gamma \\ \boldsymbol{\alpha}_\gamma \end{bmatrix} = \mathbf{X}_\gamma \beta_\gamma + \mathbf{Z}_\gamma \boldsymbol{\alpha}_\gamma$$

and

$$\mathbf{I}_\delta \boldsymbol{\delta} = \mathbf{U}_\delta \mathbf{U}'_\delta \boldsymbol{\delta} = [\mathbf{U}_{\delta n} : \mathbf{U}_{\delta r}] \begin{bmatrix} \mathbf{U}'_{\delta n} \\ \mathbf{U}'_{\delta r} \end{bmatrix}$$
$$\boldsymbol{\delta} = [\mathbf{X}_\delta : \mathbf{Z}_\delta] \begin{bmatrix} \boldsymbol{\beta}_\delta \\ \boldsymbol{\alpha}_\delta \end{bmatrix} = \mathbf{X}_\delta \boldsymbol{\beta}_\delta + \mathbf{Z}_\delta \boldsymbol{\alpha}_\delta.$$

Table 1 gives expressions for $\mathbf{X}_\gamma$, β_γ, $\mathbf{Z}_\gamma$, $\boldsymbol{\alpha}_\gamma$, $\mathbf{X}_\delta$, $\boldsymbol{\beta}_\delta$, $\mathbf{Z}_\delta$, and $\boldsymbol{\alpha}_\delta$ when the temporal random effect follows a RW1 and a RW2.

Consequently, if the temporal random effect follows a RW1, Model (8) can be expressed as

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + \mathbf{1}_{TS}\beta_\xi + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + \mathbf{1}_{TS}\beta_\gamma + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma + [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta, \tag{9}$$

where $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma^2_\xi (\lambda_\xi \widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{\xi-1})^{-1}\right)$, $\mathbf{I}_{\xi-1}$ is an identity matrix of dimension $(S-1)$, $\boldsymbol{\alpha}_\gamma \sim N\left(\mathbf{0}, \sigma^2_\gamma \widetilde{\boldsymbol{\Sigma}}^{-1}_\gamma\right)$, and $\boldsymbol{\alpha}_\delta \sim N\left(\mathbf{0}, \sigma^2_\delta \widetilde{\boldsymbol{\Sigma}}^{-1}_\delta\right)$. If we remove the repeated columns $\mathbf{1}_{TS}$ (corresponding to β_ξ, β_γ, and $\boldsymbol{\beta}_\delta$), $\mathbf{1}_T \otimes \mathbf{U}_{\xi r}$, and $\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$ (corresponding to $\boldsymbol{\beta}_\delta$), this leaves the following model

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta. \tag{10}$$

Removing the repeated columns leads to the linear constraints $\sum_{i=1}^S \xi_i = 0$, $\sum_{t=1}^T \gamma_t = 0$, $\sum_{i=1}^S \delta_{it} = 0$, $\forall t$ and $\sum_{t=1}^T \delta_{it} = 0$, $\forall i$. Note that if the ICAR prior is considered

Table 1 Expressions for the reparameterized temporal and spatio-temporal random effects for RW1 and RW2 temporal priors

Design matrices	Coefficients
RW1	
$\mathbf{X}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma n} = \mathbf{1}_{TS}$	$\beta_\gamma = \mathbf{U}'_{\gamma n}\boldsymbol{\gamma} = \mathbf{1}'_T\boldsymbol{\gamma}$
$\mathbf{Z}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma r} = \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$	$\boldsymbol{\alpha}_\gamma = \mathbf{U}'_{\gamma r}\boldsymbol{\gamma}$,
$\mathbf{X}_\delta = \mathbf{U}_{\delta n} = [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S]$	$\boldsymbol{\beta}_\delta = \mathbf{U}'_{\delta n}\boldsymbol{\delta}$
$\mathbf{Z}_\delta = \mathbf{U}_{\delta r} = [\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r}]$	$\boldsymbol{\alpha}_\delta = \mathbf{U}'_{\delta r}\boldsymbol{\delta}$
RW2	
$\mathbf{X}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma n} = [\mathbf{1}_{TS} : \mathbf{t}^* \otimes \mathbf{1}_S]$	$\boldsymbol{\beta}_\gamma = \mathbf{U}'_{\gamma n}\boldsymbol{\gamma} = [\mathbf{1}_T : \mathbf{t}^*]'\boldsymbol{\gamma}$
$\mathbf{Z}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma r} = \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$	$\boldsymbol{\alpha}_\gamma = \mathbf{U}'_{\gamma r}\boldsymbol{\gamma}$
$\mathbf{X}_\delta = \mathbf{U}_{\delta n} = [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{U}_{\xi r}]$	$\boldsymbol{\beta}_\delta = \mathbf{U}'_{\delta n}\boldsymbol{\delta}$
$\mathbf{Z}_\delta = \mathbf{U}_{\delta r} = [\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r}]$	$\boldsymbol{\alpha}_\delta = \mathbf{U}'_{\delta r}\boldsymbol{\delta}$

for the spatial random effect, then $(\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi}$ is transformed into $\mathbf{1}_{TS}\boldsymbol{\beta}_\xi + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi$, and the identifiable model takes the same form as Model (10), where now $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right)$.

If instead the temporal random effect follows a RW2, Model (8) can be expressed as

$$\begin{aligned}\log \mathbf{r} = \mathbf{1}_{TS}\eta + \mathbf{1}_{TS}\beta_\xi + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + [\mathbf{1}_{TS} : \mathbf{t}^* \otimes \mathbf{1}_S]\boldsymbol{\beta}_\gamma \\ + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma + [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S : \mathbf{t}^* \\ \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{U}_{\xi r}]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta.\end{aligned} \tag{11}$$

where $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma_\xi^2(\lambda_\xi \widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{\xi-1})^{-1}\right)$, $\boldsymbol{\alpha}_\gamma \sim N\left(\mathbf{0}, \sigma_\gamma^2 \widetilde{\boldsymbol{\Sigma}}_\gamma^{-1}\right)$, and $\boldsymbol{\alpha}_\delta \sim N\left(\mathbf{0}, \sigma_\delta^2 \widetilde{\boldsymbol{\Sigma}}_\delta^{-1}\right)$. If we remove the repeated columns $\mathbf{1}_{TS}$ (corresponding to $\boldsymbol{\beta}_\xi$, $\boldsymbol{\beta}_\gamma$, and $\boldsymbol{\beta}_\delta$), $\mathbf{1}_T \otimes \mathbf{U}_{\xi r}$, $\mathbf{t}^* \otimes \mathbf{1}_S$ and $\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$ (corresponding to $\boldsymbol{\beta}_\delta$), this leaves the following model

$$\begin{aligned}\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + (\mathbf{t}^* \otimes \mathbf{1}_S)\boldsymbol{\beta}_\gamma + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma \\ + (\mathbf{t}^* \otimes \mathbf{U}_{\xi r})\boldsymbol{\beta}_\delta + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta.\end{aligned} \tag{12}$$

If the ICAR prior were considered for the spatial random effect the identifiable model would be (12), but then $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right)$. The equivalent linear constraints are the same as in the RW1 case because the model does not include a linear trend, so that no additional identifiability issue arises. See "Appendix 1" for more details.

It is important to highlight that in Models (9) and (11) we have deleted the repeated terms $(\mathbf{1}_T \otimes \mathbf{U}_{\xi r})$ and $(\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)$ from the fixed effects arising from the reparameterization of the interaction random effect $\boldsymbol{\delta}$. We could have deleted the same terms in the reparameterization of the main spatial and temporal random effect $\boldsymbol{\xi}$ and $\boldsymbol{\gamma}$ respectively, but if we did this, it would imply that the spatial and temporal main effects were fixed instead of random, as $(\mathbf{1}_T \otimes \mathbf{U}_{\xi r})$ and $(\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)$ would only appear in the fixed part arising from the reparameterization of the interaction random effect $\boldsymbol{\delta}$. In particular, treating the spatial and temporal main effects as fixed effects would imply that they would not be smoothed at all.

4 Model fitting

Model fitting and inference with spatial and spatio-temporal disease mapping models have usually been done using either an empirical Bayes (EB) or fully Bayes (FB) approach. In the EB approach, penalized quasi-likelihood (PQL) has been widely used (see for example MacNab and Dean 2001; Dean et al. 2004; Ugarte et al. 2008, 2009, 2010, 2012). From a FB perspective, Markov chain Monte Carlo (McMC) techniques have been used because the posterior distributions usually cannot be obtained in closed form (see, for example Bernardinelli et al. 1995; Knorr-Held and Besag 1998; Knorr-Held 2000; Best et al. 2005; Ainsworth and Dean 2006; Martínez-Beneito et al. 2008; Ugarte et al. 2009). Although these techniques have been widely used, the implementation may not be easy for practitioners as algorithms have to be carefully chosen (Knorr-Held and Rue 2002; Schmid and Held 2004), and difficulties such as long computing times and large Monte Carlo errors are usually present for complex models (Schrödle et al. 2011). An alternative to McMC for Bayesian inference based on integrated nested Laplace approximations, known as INLA, has been recently proposed (Rue et al. 2009). This technique can be used easily in the free software `R` using the package `R-INLA`. In this section we show how the PQL technique naturally applies linear constraints and we provide guidelines for including constraints in INLA. Linear constraints are easy to deal with when the temporal random effect follows RW1. If the distribution of the temporal random effect is RW2, however, constraints are a bit more difficult to specify and place.

Consider the spatio-temporal Model (8). PQL requires a working vector and the restricted maximum likelihood Eqs. (Harville 1977). The components of the working vector are

$$\mathbf{Y}^* = \mathbf{X}\eta + \mathbf{Z}_1\boldsymbol{\xi} + \mathbf{Z}_2\boldsymbol{\gamma} + \mathbf{Z}_3\boldsymbol{\delta} + (\mathbf{Y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}),$$

where $\mathbf{X}$ is the fixed effects matrix (here a column of ones), $\mathbf{Z}_1 = \mathbf{1}_T \otimes \mathbf{I}_\xi$ is the design matrix of the main spatial random effect $\boldsymbol{\xi}$, $\mathbf{Z}_2 = \mathbf{I}_\gamma \otimes \mathbf{1}_S$ is the design matrix of the main temporal random effect $\boldsymbol{\gamma}$, $\mathbf{Z}_3 = \mathbf{I}_\delta$ is the design matrix of the interaction term $\boldsymbol{\delta}$, $\boldsymbol{\mu}$ is the vector of means of the Poisson distribution, g is the link function (here the logarithmic function), and $g'(\boldsymbol{\mu}) = 1/\boldsymbol{\mu}$. Then a correspondence with a normal mixed model is attained as

$$\mathbf{Y}^* = \mathbf{X}\eta + \mathbf{Z}_1\boldsymbol{\xi} + \mathbf{Z}_2\boldsymbol{\gamma} + \mathbf{Z}_3\boldsymbol{\delta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} = (\mathbf{Y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{W}^{-1})$, and $\mathbf{W} = diag(\mu_{it})$. The fixed effect estimator is obtained as $\hat{\eta} = \left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}^*$, where $\mathbf{V} = \mathbf{W}^{-1} + \mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{G}_2\mathbf{Z}_2' + \mathbf{Z}_3\mathbf{G}_3\mathbf{Z}_3'$, and $\mathbf{G}_1 = \sigma_\xi^2\mathbf{D}^{-1}$ or $\mathbf{G}_1 = \sigma_\xi^2\mathbf{Q}_\xi^-$ depending on whether the spatial effect follows the LCAR prior (2) or the ICAR prior respectively, $\mathbf{G}_2 = \sigma_\gamma^2\mathbf{Q}_\gamma^-$, and $\mathbf{G}_3 = \sigma_\delta^2\mathbf{Q}_\delta^-$ (see for example Ugarte et al. 2010 for details). The random effects are predicted as

$$\hat{\boldsymbol{\xi}} = \hat{\mathbf{G}}_1\mathbf{Z}_1'\hat{\mathbf{V}}^{-1}(\mathbf{Y}^* - \mathbf{X}\hat{\eta}),$$
$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}_2\mathbf{Z}_2'\hat{\mathbf{V}}^{-1}(\mathbf{Y}^* - \mathbf{X}\hat{\eta}),$$
$$\hat{\boldsymbol{\delta}} = \hat{\mathbf{G}}_3\mathbf{Z}_3'\hat{\mathbf{V}}^{-1}(\mathbf{Y}^* - \mathbf{X}\hat{\eta}).$$

If $\mathbf{G}_1 = \sigma_\xi^2 \mathbf{Q}_\xi^-$, then, the PQL technique automatically imposes the usual sum-to-zero constraint $\sum_{i=1}^S \xi_i = 0$. This is clear as $\mathbf{Q}_\xi \mathbf{1}_S = \mathbf{0}$, and hence $\sum_{i=1}^S \hat{\xi}_i = \hat{\sigma}_\xi^2 \Big(\mathbf{Q}_\xi^- \mathbf{Z}_1' \hat{\mathbf{V}}^{-1} (\mathbf{Y}^* - \mathbf{X}\hat{\eta})\Big)' \mathbf{1}_S = \hat{\sigma}_\xi^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_1 \mathbf{Q}_\xi^- \mathbf{1}_S = 0$. Note that if $\mathbf{x}$ is an eigenvector of $\mathbf{Q}_\xi$ that has zero eigenvalue, then $\mathbf{x}$ is an eigenvector of $\mathbf{Q}_\xi^-$ that has zero eigenvalue (see for example Harville 2008, chap. 21, p. 546). Consequently, $\mathbf{Q}_\xi^- \mathbf{1}_S = \mathbf{0}$. Similarly, the sum-to-zero constraint is automatically imposed for the temporal random effects. Furthermore, if the LCAR prior is used for the spatial effect, i.e., if $\mathbf{G}_1 = \sigma_\xi^2 \mathbf{D}^{-1} = \sigma_\xi^2 (\lambda_\xi \mathbf{Q}_\xi + (1-\lambda_\xi)\mathbf{I}_\xi)^{-1}$, the implicit intercept, $\sum_{i=1}^S \xi_i$, is estimated as zero and hence PQL automatically imposes the sum-to-zero constraint. Note that in general, if $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ that has non-zero eigenvalue λ, then $\mathbf{x}$ is an eigenvector of $\mathbf{A}^{-1}$ that has non-zero eigenvalue $1/\lambda$ (see for example Harville 2008 chap 21, p. 527). Consequently, as $\mathbf{Q}_\xi \mathbf{1}_S = \mathbf{0}$ and $\mathbf{I}_\xi \mathbf{1}_S = \mathbf{1}_S$, $\mathbf{D}\mathbf{1}_S = (\lambda_\xi \mathbf{Q}_\xi + (1-\lambda_\xi)\mathbf{I}_\xi)\mathbf{1}_S = (1-\lambda_\xi)\mathbf{1}_S$, so that $\mathbf{D}^{-1}\mathbf{1}_S = \frac{1}{(1-\lambda_\xi)}\mathbf{1}_S$. Hence, $\sum_{i=1}^S \hat{\xi}_i = \hat{\sigma}_\xi^2 \big(\hat{\mathbf{D}}^{-1} \mathbf{Z}_1' \hat{\mathbf{V}}^{-1} (\mathbf{Y}^* - \mathbf{X}\hat{\eta})\big)' \mathbf{1}_S = \hat{\sigma}_\xi^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_1 \hat{\mathbf{D}}^{-1} \mathbf{1}_S = \frac{\hat{\sigma}_\xi^2}{(1-\hat{\lambda}_\xi)} (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{1}_{TS} = 0$ taking into account that

$$
\begin{aligned}
(\mathbf{Y}^* - \mathbf{X}\hat{\eta})'\hat{\mathbf{V}}^{-1}\mathbf{X} &= \left(\mathbf{Y}^* - \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}^*\right)'\hat{\mathbf{V}}^{-1}\mathbf{X} = \\
&= \mathbf{Y}^{*\prime}\left(\mathbf{I} - \hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)\hat{\mathbf{V}}^{-1}\mathbf{X} = \\
&= \mathbf{Y}^{*\prime}\left(\hat{\mathbf{V}}^{-1}\mathbf{X} - \hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right) = \mathbf{0}.
\end{aligned}
$$

Then if the intercept is included in the model, the vector $\mathbf{1}_{TS}$ is one of the columns of $\mathbf{X}$ and therefore $(\mathbf{Y}^* - \mathbf{X}\hat{\eta})'\hat{\mathbf{V}}^{-1}\mathbf{1}_{TS} = 0$.

PQL also automatically imposes the constraints $\sum_{i=1}^S \delta_{it} = 0 \quad \forall t$ and $\sum_{t=1}^T \delta_{it} = 0 \quad \forall i$ if the temporal random effect follows a RW1. This is also easy to see. If we define $\mathbf{e}_i$, $i = 1 \ldots, S$, to be a vector of length S with a one in the i-th position and zero elsewhere and $\mathbf{u}_t$, $t = 1 \ldots, T$, to be a vector of length T with a one in the t-th position and zero elsewhere, then

$$
\begin{aligned}
\sum_{i=1}^S \hat{\delta}_{it} &= \hat{\sigma}_\delta^2 \big(\mathbf{Q}_\delta^- \mathbf{Z}_3' \hat{\mathbf{V}}^{-1} (\mathbf{Y}^* - \mathbf{X}\hat{\eta})\big)' (\mathbf{u}_t \otimes \mathbf{1}_S) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 \mathbf{Q}_\delta^- (\mathbf{u}_t \otimes \mathbf{1}_S) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^- (\mathbf{u}_t \otimes \mathbf{1}_S) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 ((\mathbf{Q}_\gamma^- \mathbf{u}_t) \otimes (\mathbf{Q}_\xi^- \mathbf{1}_S)) = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{t=1}^T \hat{\delta}_{it} &= \hat{\sigma}_\delta^2 \big(\mathbf{Q}_\delta^- \mathbf{Z}_3' \hat{\mathbf{V}}^{-1} (\mathbf{Y}^* - \mathbf{X}\hat{\eta})\big)' (\mathbf{1}_T \otimes \mathbf{e}_i) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 \mathbf{Q}_\delta^- (\mathbf{1}_T \otimes \mathbf{e}_i) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^- (\mathbf{1}_T \otimes \mathbf{e}_i) \\
&= \hat{\sigma}_\delta^2 (\mathbf{Y}^* - \mathbf{X}\hat{\eta})' \hat{\mathbf{V}}^{-1} \mathbf{Z}_3 ((\mathbf{Q}_\gamma^- \mathbf{1}_T) \otimes (\mathbf{Q}_\xi^- \mathbf{e}_i)) = 0.
\end{aligned}
$$

Hence, if the distribution of the temporal random effect is RW1, PQL automatically places correct constraints. However, if the distribution of the temporal random effect is RW2, then PQL imposes more restrictions than needed unless extra fixed-effect terms—a common linear trend or linear trends for each area—are added to the model. A RW2 prior for the temporal random effect implies that PQL automatically imposes the constraint $\mathbf{Q}_\gamma \mathbf{t}^* = \mathbf{0}$, so that $\sum_{t=1}^T t\hat{\gamma}_t = 0$, but this constraint is not needed unless a linear trend is in the model as a fixed effect. Similarly $(\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)(\mathbf{t}^* \otimes \mathbf{e}_i) = \mathbf{0}$, so that PQL automatically imposes the constraint $\sum_{t=1}^T t\hat{\delta}_{it} = 0, \forall i$. Consequently, PQL places more constraints than are needed unless we explicitly consider the model reformulation given in Eq. (12). Using the reformulated Model (10) instead of Model (8) also avoids the previously mentioned inconvenience of variance inflation. The reason is that when we use the LCAR prior for the spatial effect, the eigenvalue corresponding to the eigenvector $\mathbf{1}_S$ is not null and hence it contributes to the variance of the intercept (see "Appendix 3"). This does not happen if the ICAR prior is considered because the eigenvalue associated with the eigenvector $\mathbf{1}_S$ is equal to zero, so that it does not contribute to the variance of the intercept. When we consider the reformulated model, the redundant intercepts disappear from the model and the variance is not inflated.

Recently, a new approximate technique called INLA, based on integrated nested Laplace approximations, has been proposed for Bayesian inference in models using latent Gaussian Markov random fields (Rue et al. 2009), which includes the models described in this paper. An attractive feature of INLA is that it can easily be used in the free software `R` (R Core Team 2016), with the package `R-INLA` (Martino and Rue 2009). `R` code to fit some of these models in INLA can be found in Ugarte et al. (2014). Details about how to place constraints in disease mapping models using INLA can be found in Schrödle and Held (2011). We recommend a careful reading of this paper to avoid misunderstandings. According to the authors, " ...the identifiability of $\boldsymbol{\delta}$ can be ensured by computing the null space of the respective structure matrix $\mathbf{R}$ and using the obtained eigenvectors as linear constraints for the estimation of $\boldsymbol{\delta}$. As a consequence, the number of linear constraints which are necessary is *always* equal to the rank deficiency of **R** [emphasis added]". This is true if the model includes an intercept and the temporal effect is modeled as a RW1. However, if a RW2 prior is used for the temporal random effect, constraints are not in fact needed for all the eigenvectors corresponding to the null eigenvalues of the precision matrix *unless* a common linear trend and area specific linear trends are included in the model as fixed effects, as shown in Sect. 3.2. Section 5 below shows the consequences of adding these needless constraints in a model with a RW2 prior for the temporal random effect. "Appendices 1 and 2" show the appropriate constraints, and "Appendix 4" provides R code to fit the model with the different types of interactions.

5 Illustration

This section uses female breast cancer mortality data (ICD-10 code 50) in Spanish provinces during the period 1990-2010 to illustrate how estimates can change if unnecessary linear constraints are unintentionally included in the model. The models are fitted in an EB approach using PQL and in a FB approach using INLA. In all of the models, a global intercept η has been included in estimating the log-risks, so we must center the spatial random effects by including the constraint $\sum_{i=1}^{S} \xi_i = 0$. Regarding estimation of the intercept using PQL, if a RW1 prior is used for time and the LCAR prior is used for the spatial random effect, the estimated standard error of the intercept is inflated when we fit the complete Model (8) with appropriate sum-to-zero constraints. In this case, $\hat{\eta} = -0.034$ and $s.e.(\hat{\eta}) = 0.036$. However, if we fit the reparameterized Model (10), $\hat{\eta} = -0.034$ and $s.e.(\hat{\eta}) = 0.0039$. If an ICAR prior is used for the spatial random effect, the same estimates are obtained from the full Model (8) with appropriate sum-to-zero constraints and from the reparameterized Model (10), i.e., the estimate and the standard error for the intercept are $\hat{\eta} = -0.034$ and $s.e.(\hat{\eta}) = 0.0039$ respectively. Using INLA (placing uniform distributions on the standard deviations, and a full Laplace strategy), the estimate of the intercept ($\hat{\eta} = -0.035$) and its standard error ($s.e.(\hat{\eta}) = 0.004$) are identical for the complete Model (8) and the reduced Model (10) when we use the ICAR for the spatial random effect. If we use the LCAR prior, the estimates of the intercept and its s.e. without constraints are $\hat{\eta} = -0.035$ and $s.e.(\hat{\eta}) = 0.048$, whereas the estimates are $\hat{\eta} = -0.035$ and $s.e.(\hat{\eta}) = 0.004$ using the constraints or the reparameterized model.

We now focus on the estimated temporal pattern $\hat{\gamma}_t$ common to all small areas. As shown in Sect. 4, if a RW2 prior is used for the temporal random effect, PQL automatically sets the constraints $\sum_{t=1}^{T} \gamma_t = 0$ and $\sum_{t=1}^{T} t\gamma_t = 0$, the latter being unnecessary because no explicit linear trend is included in the model as a fixed effect. Figure 1 on the upper left shows the estimated temporal patterns obtained with a RW1 and a RW2 using PQL with Model (8). If we compare the fits, we see a different trend for the RW2 due to the unnecessary constraint $\sum_{t=1}^{T} t\gamma_t = 0$, instead of the expected smoother version of the RW1 fit. However, if the reparameterized Models (10) and (12) are fitted (upper right panel in Fig. 1), where the repeated (or linearly dependent) columns of the fixed effect matrix are deleted, the results of the two fits are much more consistent. Because only the repeated columns are removed, the unnecessary constraint $\sum_{t=1}^{T} t\gamma_t = 0$ is not placed on the fit. If the models are fitted using INLA, once the appropriate constraints are specified for the temporal effect γ_t, the temporal pattern estimated for the original Model (8) (Fig. 1, bottom left) and the reparameterized Models (10) and (12) (Fig. 1, bottom right) are almost identical, indicating that INLA does not inherently place unnecessary constraints on the temporal effects.

Now consider the estimates of the spatio-temporal random effect $\boldsymbol{\delta} \sim N\left(\mathbf{0}, \sigma_\delta^2(\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^-\right)$. If a RW2 prior is considered for the temporal random effect and the model is fitted using PQL, the following linear constraints are automatically imposed: $\sum_{i=1}^{S} \delta_{it} = 0, \forall t$, $\sum_{t=1}^{T} \delta_{it} = 0, \forall i$, and $\sum_{t=1}^{T} t\delta_{it} = 0, \forall i$. The latter constraints are unnecessary and force the fitted area-specific risk evolution to have a very restricted shape. Figure 2 shows the estimated interaction effects for three selected provinces using PQL. The top row in Fig. 2 shows the estimates when both RW1 and RW2 are considered in the original Model (8). Clearly the unnecessary constraints for a RW2 make the fit very different from that obtained with a RW1: the linear component of the RW1 fit is absent in the RW2 fit. The bottom row in Fig. 2 shows the estimates using the reparameterized Models (10) and (12). The results seem to be more sensible because the RW1 and RW2 models give similar fits, with the RW2 fits being smoother as expected.

The results obtained with INLA are also interesting. Proceeding as suggested in Schrödle and Held (2011), with all eigenvectors in the null space of $\mathbf{Q}_\delta = \mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi$ used as constraints to identify $\boldsymbol{\delta}$, when the RW2 prior is used the interaction fit given by INLA is very similar to the restricted interaction fit using PQL (top row of Fig. 3). This problem is solved if the reparameterized Models (10)

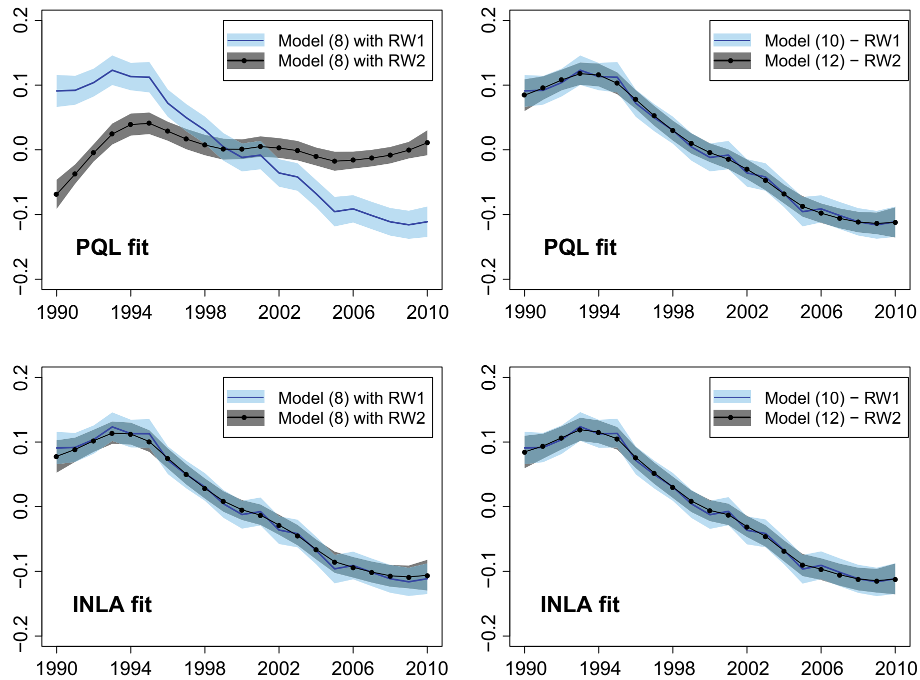

Fig. 1 Estimated temporal trend $\hat{\gamma}_t$ with PQL (*top*) and INLA (*bottom*), and corresponding confidence bands. *Blue bands* refers to RW1 and *grey bands* are for RW2

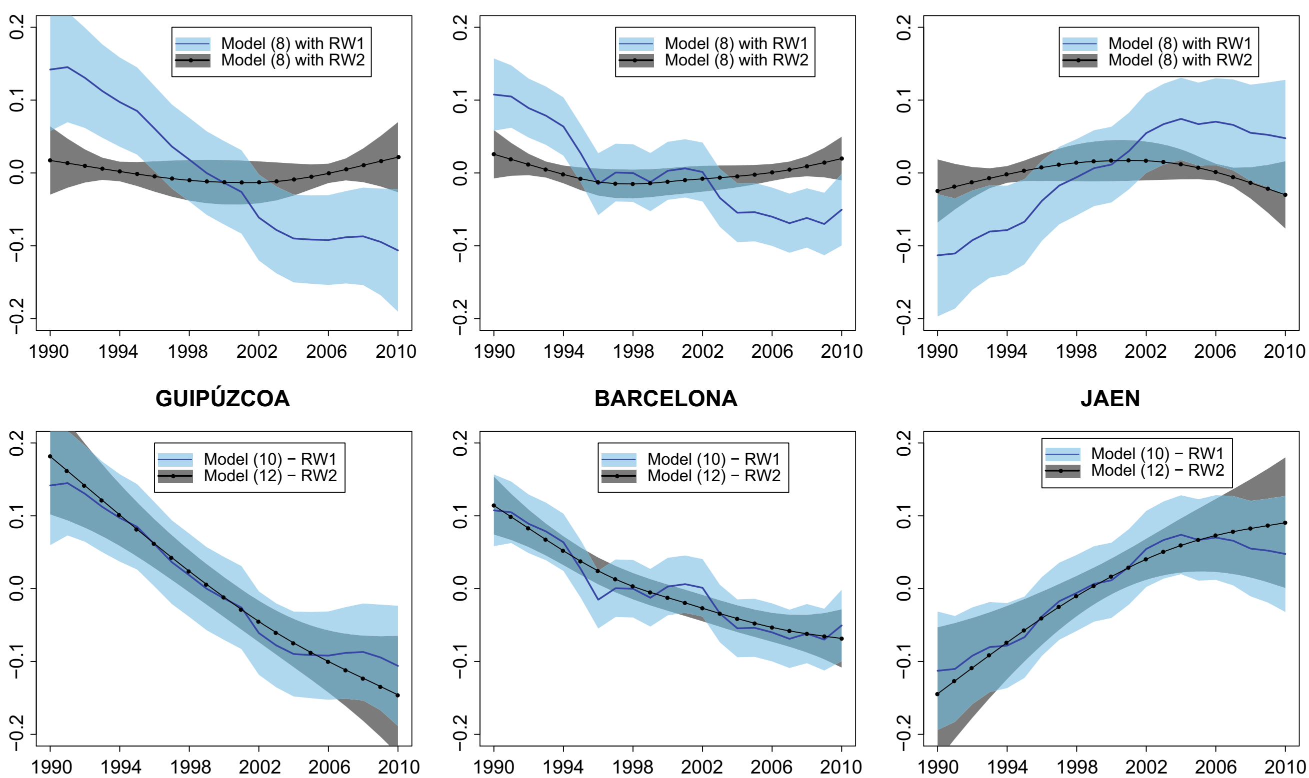

Fig. 2 Space-time interaction random effect $\hat{\delta}_{it}$ estimated with PQL and corresponding confidence bands. *Blue bands* refers to RW1 and *grey bands* are for RW2

and (12) are used instead (bottom row of Fig. 3). If instead of reparameterizing the model, we fit the original model using INLA and apply only the appropriate sum-to-zero constraints to $\boldsymbol{\delta}$, the resulting fit is similar to those from the re-parameterized models with slight differences in some areas with low populations. Figure 4 shows the

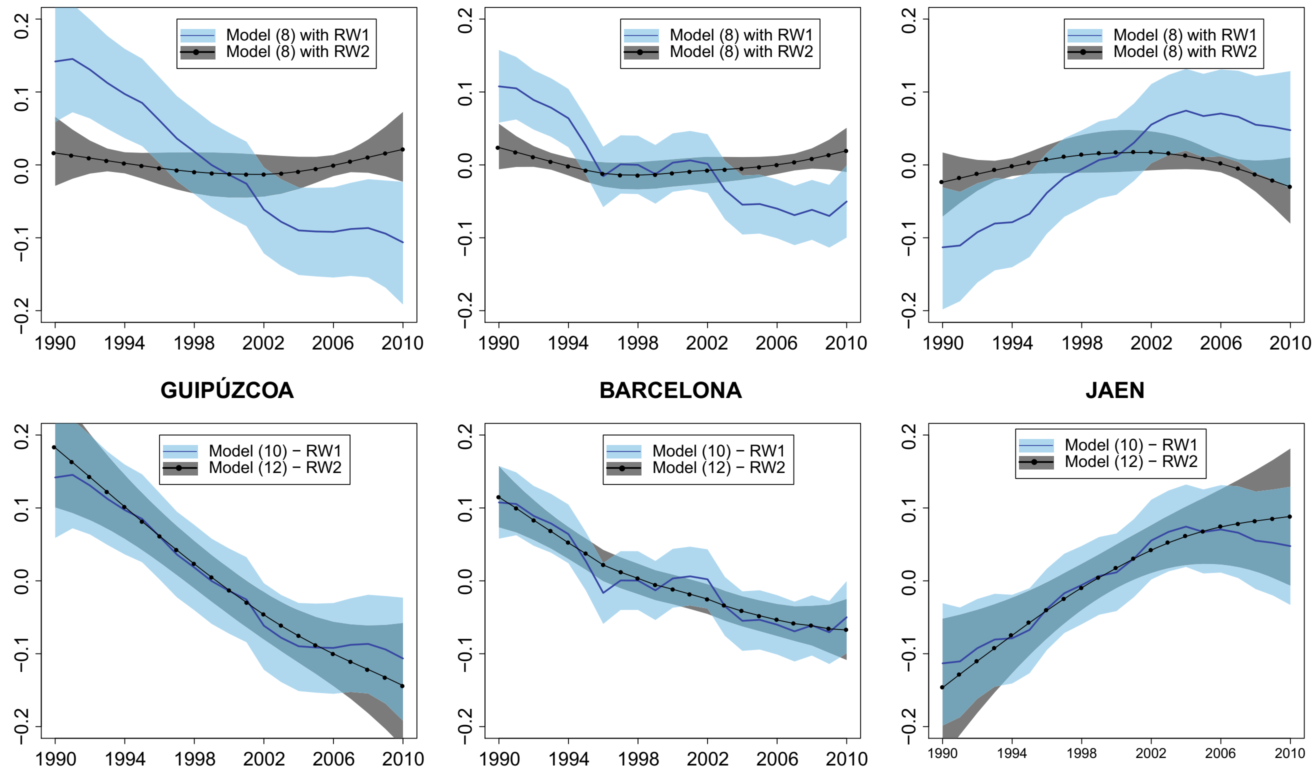

Fig. 3 Space-time interaction random effect $\hat{\delta}_{it}$ estimated with INLA and corresponding confidence bands. *Blue bands* refers to RW1 and *grey bands* are for RW2

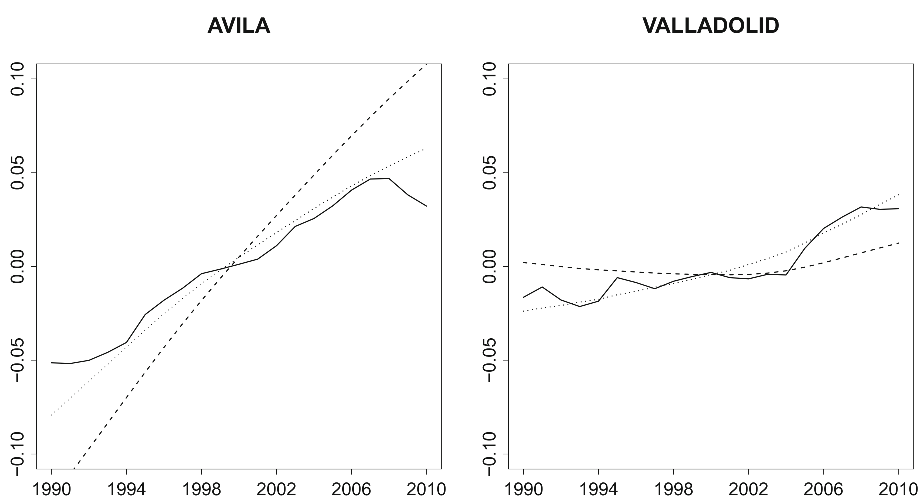

Fig. 4 Space-time interaction random effect $\hat{\delta}_{it}$ estimated with INLA for Model (8) with RW1 (*solid line*), Model (8) with RW2 and appropriate sum-to-zero constraints (*dashed line*) and the reparameterized Model (12) (*dotted line*)

estimated interaction random effect in two low-population provinces when fitting Model (8) using INLA with a RW1 (solid line), with a RW2 and appropriate sum-to-zero constraints (dashed line), and fitting the reparameterized Model (12) (dotted line). The estimates with the correct constraints and with the reparameterized model differ, the latter being similar to estimates using PQL. However, these differences do not have a great impact on the final risk estimates.

Finally, to see the effects of unnecessary constraints on the final risk estimates, Fig. 5 displays the estimated relative risks with PQL and INLA for two provinces, Guipuzcoa (left column) and Barcelona (right column). A dashed horizontal line representing a standard mortality

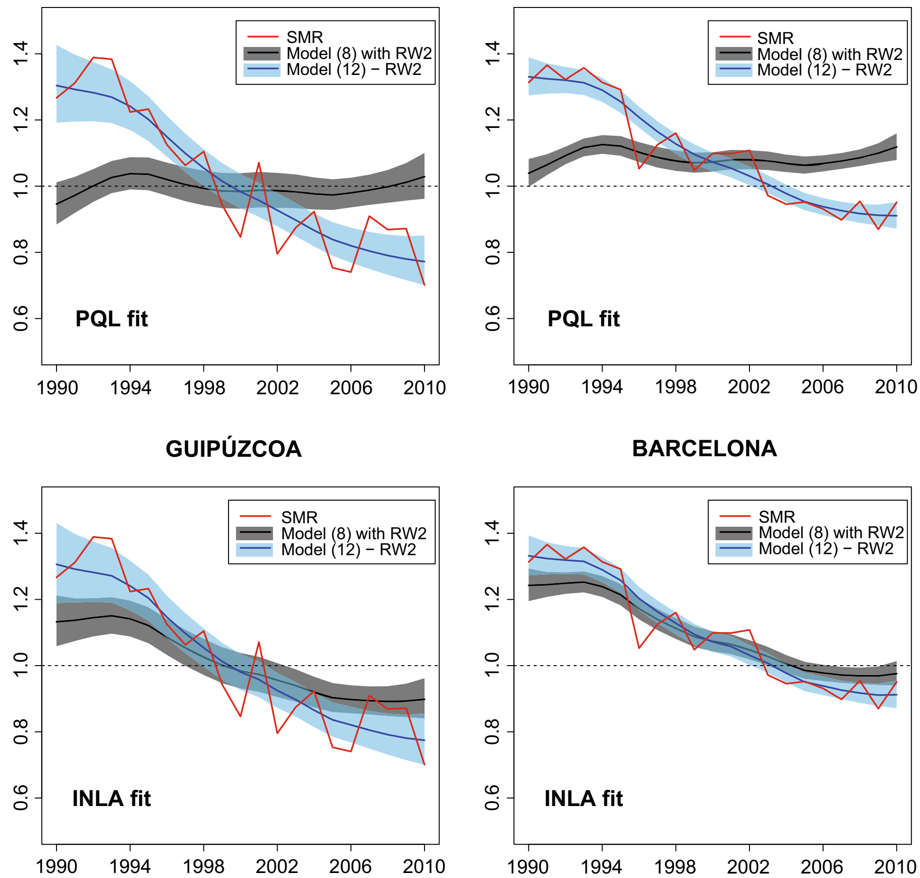

Fig. 5 Standardized mortality ratios (*red lines*), and estimated relative risks with PQL (*top row*) and INLA (*bottom row*) for the provinces of Guipuzcoa (*left column*) and Barcelona (*right column*). *Black lines* and *grey confidence bands* correspond to Model (8) with unnecessary constraints, and *blue lines* and *blue confidence bands* represent the estimates from the reparameterized Model (12). The *dashed line* is a reference line to compare provinces risks with the whole of Spain

ratio of 1 is used as a reference to compare provinces with the whole country. Figure 5's top row shows the PQL fit; its bottom row shows the INLA fit. If Model (8) is fitted using PQL, the estimated relative risk (black continuous line with grey confidence band) do not track the SMRs' trend (red line). On the other hand, if the reparameterized Model (12) is fitted, the estimated relative risks (blue line with blue confidence band) track the SMRs very well. The confidence band for the estimated relative risks has been constructed using the estimated mean squared error of the logrisk and the delta method (see Ugarte et al. 2008, 2010; Adin et al. 2016). The same effect is also observed in the INLA fit (bottom row). If Model (8) is fitted using as many constraints in the interaction as eigenvectors spanning the null space of the interaction covariance matrix, the estimated relative risks (black continuous line with grey credibility band) are wrong. However, if Model (8) with appropriate constraints or the reparameterized Model (12) is fitted, the estimated relative risks (blue line with blue credibility band) are correct. It should be noticed that PQL places unnecessary constraints on the temporal main effect and on the interaction effect, whereas using INLA, the user must explicitly place the correct constraints on the interaction term. Consequently, the effect of the unnecessary constraints is stronger in the PQL fit. Nevertheless, the effect of unnecessary constraints in the final risk is serious enough to be taken into account.

To compare models with constraints versus reparameterized models, we computed AIC (PQL fitting) and DIC (INLA fitting). INLA does not take into account the deviance of the saturated model to compute the DIC, so we have computed the AIC without this term to make them comparable. More precisely, the deviance of the saturated model is

$$2\log\left(\prod \frac{e^{-Y_{it}}Y_{it}^{Y_{it}}}{Y_{it}!}\right) = 2\sum(-Y_{it} + Y_{it}\log(Y_{it}) - \log(Y_{it}!)).$$

Table 2 shows AIC for models fit using PQL adjusted as just noted and DIC for models fit using INLA. Regarding PQL, the two models have similar AICs when the temporal distribution is a RW1, showing equivalent results using constraints in the original model or fitting the reparameterized model. However, when the temporal random effect is a RW2, a higher AIC is observed for the original model, for which PQL imposes more constraints than needed. Results with INLA are similar to those from PQL. For RW1, the recommended constraints are appropriate, but for RW2 using more constraints than needed leads to poor fits. Table 2 uses "Model 8*" to denote Model (8) fitted in INLA with the appropriate sum-to-zero constraints. The DICs obtained with Model 8* are slightly higher than the analogous DICs from the reparameterized model. This may reflect the effect shown

Table 2 Model comparisons in the analysis of female breast cancer mortality data in Spain. Model 8 is the original model, Models 10 and 12 are the reparameterized models, and Model 8* denotes Model (8) fitted in INLA with the appropriate constraints. For the INLA fit, a full Laplace strategy has been used, and $\bar{D}$ and p_D are the mean deviance and the effective number of parameters respectively. The computing time is given in seconds

	PQL							
	RW1				RW2			
	Deviance	Df	AIC	Time	Deviance	Df	AIC	Time
ICAR models								
Model 8	7410.9	145.6	7702.1	75	8802.3	66.1	8934.6	100
Model 10/12	7411.8	145.2	7702.1	240	7494.5	110.8	7716.1	400
LCAR models								
Model 8	7409.9	146.0	7701.8	95	8801.3	66.5	8934.3	125
Model 10/12	7411.1	145.4	7701.8	170	7493.7	111.1	7716.0	240
	INLA							
	$\bar{D}$	p_D	DIC	Time	$\bar{D}$	p_D	DIC	Time
ICAR models								
Model 8	7555.0	149.7	7704.8	60	7798.1	71.5	7869.6	120
Model 8*					7601.9	120.4	7722.4	55
Model 10/12	7555.0	149.7	7704.8	1030	7601.7	117.2	7718.9	705
LCAR models								
Model 8	7554.2	150.3	7704.5	100	7797.5	71.7	7869.2	235
Model 8*					7601.2	120.9	7722.1	95
Model 10/12	7554.2	150.3	7704.5	1640	7600.9	117.9	7718.8	1215

in Fig. 4. Finally, the reason AIC and DIC are very different for the RW2 is that PQL places incorrect constraints on both the temporal main effects γ_t and the interaction effects δ_{it}, while in the INLA fit we placed correct constraints on the main effects but incorrect ones on the interaction effects, where we followed Schrödle and Held (2011)'s recommendation to use as many constraints as the rank deficiency of the precision matrix, which as noted is not correct. Table 2 also shows computing time (in seconds) required to fit the models. The PQL fits were run on a LENOVO personal computer with a 3.1 GHz Intel Core i5 processor and 6GB RAM using R (version 3.2.2). The INLA fits were run on a twin super-server with four processors, Intel Xeon 6C and 96GB RAM, using R (version 3.2.2) and the R package INLA (version 0.0-1455098891, dated 2016-02-10). Computing time is higher for the reparameterized models. Consequently, if INLA is used to fit the spatio-temporal Model (8) with a RW2 prior for the temporal component, it is recommended to place appropriate constraints instead of reparameterizing the model.

6 Discussion

Statistical models used in spatial and spatio-temporal disease mapping have become more and more sophisticated to allow proper analyses of real data. This complexity has brought some challenges, model identifiability being one of the most important. There are plenty of papers on spatial and spatio-temporal disease mapping; most consider sum-to-zero constraints to achieve identifiability but do not clearly establish why and how the constraints should be imposed.

This paper's main objective was to clarify this issue, providing practical guidelines when spatio-temporal disease mapping models are fitted using PQL or INLA. In both approaches (empirical or fully Bayes), one of the more widely used priors for spatial random effects is the intrinsic CAR (ICAR). Recently, certain disadvantages of this prior have been reported, for example, it produces negative correlations among regions located far apart. These limitations have led some authors to use the Leroux prior (LCAR) as a possible alternative to the ICAR. The LCAR prior does not produce such negative correlations and has the advantage of including a parameter that quantifies spatial dependence as well as unstructured heterogeneity. However, if the LCAR prior is used and Model (8) (a spatio-temporal model with a RW1 prior for the main temporal effect) is fitted using PQL, an undesired variance inflation of the intercept estimate occurs even if adequate restrictions on the spatial effects are used. In addition, if a RW2 prior is used for the temporal random effects, PQL automatically places unnecessary constraints on the estimates, which leads to an erroneous estimate of the temporal trend. Both problems can be fixed with a reparameterization of the model based on the spectral decomposition of the precision matrices of the spatial, temporal, and spatio-temporal random effects, as shown in this paper. We would like to remark again the difference between the ICAR and the LCAR priors. While the ICAR prior is improper, creating an identifiability issue with the intercept, the LCAR is proper and the intercept is partially identified by its prior. However, there is never an advantage to retaining the separate intercept arising from the LCAR, whereas omitting it has the advantage of avoiding variance inflation, so it should be omitted. If additional spatially varying fixed effects are included in the model, collinearity problems with the spatial main effect could appear (see Reich et al. 2006) affecting the standard error of the estimates. However, this is beyond the scope of this paper and deserves further research. Here, our focus is on adequate restrictions in disease mapping models commonly used to produce maps in vital-statistics agencies, public health institutions, and research centers. The models in this paper are very valuable because in many diseases risk factors (predictors) are unknown, and the goal of disease mapping is to unveil geographical and temporal patterns which may help to identify potential risk factors and to elucidate the aetiology of a disease. The issues examined in this paper are also applicable to other environmental settings where random effects of this or a similar type are included.

For Bayesian inference, model fitting can be done using McMC or INLA; INLA is popular nowadays and we have focused on it. This paper showed how to specify appropriate constraints in INLA when fitting the more common spatio-temporal models, including four types of spatio-temporal interactions (Knorr-Held 2000; see "Appendices 1 and 2"). Previous papers have not clarified this aspect of using INLA in enough detail and proper constraint specification is crucial for practitioners. In particular it is *not* the case that the number of linear constraints needed to identify the interaction effect is always equal to the rank deficiency of the precision matrix (Schrödle and Held 2011). Placing more restrictions than needed again leads to erroneous estimates.

Summing up, when using PQL to fit spatio-temporal disease mapping models, our recommendation is to reparameterize the model using the spectral decomposition of the precision matrices of the random effects before fitting. If INLA is used to fit models, appropriate constraints must be identified and used. Doing so gives correct results without incurring the extra computing time required to fit the reparameterized model.

Acknowledgements This work has been supported by the Spanish Ministry of Economy and Competitiveness (project MTM2014-51992-R), and by the Health Department of the Navarre Government (Project 113, Res.2186/2014). We would like to thank the National Epidemiology Center (area of Environmental Epidemiology and Cancer) for providing the data, originally created by the Spanish Statistical Office. Thanks are also given to two anonymous reviewers for their comments that have contributed to improve the paper.

Appendix 1

Appendix 1 shows how the usual sum-to-zero constraints are derived.

Appendix 1.1

Given the spatio-temporal model of Eq. (8)

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi} + (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} + \mathbf{I}_\delta \boldsymbol{\delta},$$

where η is the log of the global risk, the spatial random effect $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma_\xi^2 \mathbf{Q}_\xi^-\right)$ follows an ICAR distribution, the temporal random effect $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \sigma_\gamma^2 \mathbf{Q}_\gamma^-\right)$ follows a RW1 distribution, and the interaction random effect $\boldsymbol{\delta} \sim N\left(\mathbf{0}, \sigma_\delta^2 (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^-\right)$ is completely structured (Type IV interaction) in space and time, the linear constraints that make this model identifiable are

$$\sum_{i=0}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \begin{aligned} &\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for } i = 1, \ldots, S \\ &\sum_{i=1}^{S} \delta_{it} = 0, \quad \text{for } t = 1, \ldots, T \end{aligned}$$

As shown in Sect. 3.2, the random effects of Eq. (8) can be reparameterized using the spectral decomposition of their covariance matrices, so this model becomes

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + [\mathbf{X}_\xi : \mathbf{Z}_\xi] \begin{bmatrix} \beta_\xi \\ \boldsymbol{\alpha}_\xi \end{bmatrix} + [\mathbf{X}_\gamma : \mathbf{Z}_\gamma] \begin{bmatrix} \beta_\gamma \\ \boldsymbol{\alpha}_\gamma \end{bmatrix} + [\mathbf{X}_\delta : \mathbf{Z}_\delta] \begin{bmatrix} \boldsymbol{\beta}_\delta \\ \boldsymbol{\alpha}_\delta \end{bmatrix}$$

where

$$\begin{aligned}
\mathbf{X}_\xi &= (\mathbf{1}_T \otimes \mathbf{I}_\xi)\mathbf{U}_{\xi n} = \mathbf{1}_{TS}, \quad \beta_\xi = \mathbf{U}'_{\xi n}\boldsymbol{\xi} = \mathbf{1}'_S\boldsymbol{\xi}, \\
\mathbf{Z}_\xi &= (\mathbf{1}_T \otimes \mathbf{I}_\xi)\mathbf{U}_{\xi r} = \mathbf{1}_T \otimes \mathbf{U}_{\xi r}, \quad \boldsymbol{\alpha}_\xi = \mathbf{U}'_{\xi r}\boldsymbol{\xi}, \\
\mathbf{X}_\gamma &= (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma n} = \mathbf{1}_{TS}, \quad \beta_\gamma = \mathbf{U}'_{\gamma n}\boldsymbol{\gamma} = \mathbf{1}'_T\boldsymbol{\gamma}, \\
\mathbf{Z}_\gamma &= (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma r} = \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S, \quad \boldsymbol{\alpha}_\gamma = \mathbf{U}'_{\gamma r}\boldsymbol{\gamma}, \\
\mathbf{X}_\delta &= \mathbf{U}_{\delta n} = [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S], \quad \boldsymbol{\beta}_\delta = \mathbf{U}'_{\delta n}\boldsymbol{\delta}, \\
\mathbf{Z}_\delta &= \mathbf{U}_{\delta r} = [\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r}], \quad \boldsymbol{\alpha}_\delta = \mathbf{U}'_{\delta r}\boldsymbol{\delta}.
\end{aligned}$$

Consequently, the model can be re-expressed as

$$\begin{aligned}
\log \mathbf{r} = {} & \mathbf{1}_{TS}\eta + \mathbf{1}_{TS}\beta_\xi + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + \mathbf{1}_{TS}\beta_\gamma + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma \\
& + [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta.
\end{aligned}$$

where $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma_\xi^2 \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right)$, $\boldsymbol{\alpha}_\gamma \sim N\left(\mathbf{0}, \sigma_\gamma^2 \widetilde{\boldsymbol{\Sigma}}_\gamma^{-1}\right)$ and $\boldsymbol{\alpha}_\delta \sim N\left(\mathbf{0}, \sigma_\delta^2 \widetilde{\boldsymbol{\Sigma}}_\delta^{-1}\right)$.

To obtain the identifiable model of Eq. (10), the repeated columns $\mathbf{1}_{TS}$, $\mathbf{1}_T \otimes \mathbf{U}_{\xi r}$ and $\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$ must be removed, which is equivalent to set $\beta_\xi = 0$, $\beta_\gamma = 0$ and $\boldsymbol{\beta}_\delta = \mathbf{0}$.

For the first two constraints, it is straightforward that

$$\beta_\xi = 0 \Longleftrightarrow \mathbf{1}'_S \boldsymbol{\xi} = 0 \Longleftrightarrow \sum_{i=1}^{S} \xi_i = 0 \quad \text{and}$$

$$\beta_\gamma = 0 \Longleftrightarrow \mathbf{1}'_T \boldsymbol{\gamma} = 0 \Longleftrightarrow \sum_{t=1}^{T} \gamma_t = 0.$$

Now decompose the constraint $\boldsymbol{\beta}_\delta = \mathbf{0}$ into its three terms:

- As in the previous case, $\mathbf{1}'_{TS}\boldsymbol{\delta} = 0 \Longleftrightarrow \sum_{i=1}^{S}\sum_{t=1}^{T} \delta_{it} = 0$.
- Denoting $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{S1}, \ldots, \delta_{1T}, \ldots, \delta_{ST})'$ it can be shown that

$$(\mathbf{1}_T \otimes \mathbf{U}_{\xi r})'\boldsymbol{\delta} = [\mathbf{U}'_{\xi r} : \cdots : \mathbf{U}'_{\xi r}]\begin{pmatrix}\delta_{11}\\ \vdots \\ \delta_{ST}\end{pmatrix}$$

$$= \mathbf{U}'_{\xi r}[\mathbf{I}_\xi : \cdots : \mathbf{I}_\xi]\begin{pmatrix}\delta_{11}\\ \vdots \\ \delta_{ST}\end{pmatrix} = \mathbf{U}'_{\xi r}\begin{pmatrix}\sum_{t=1}^{T}\delta_{1t}\\ \vdots \\ \sum_{t=1}^{T}\delta_{St}\end{pmatrix},$$

and

$$(\mathbf{1}_T \otimes \mathbf{U}_{\xi r})'\boldsymbol{\delta} = \mathbf{0} \Longleftrightarrow \mathbf{U}'_{\xi r}\begin{pmatrix}\sum_{t=1}^{T}\delta_{1t}\\ \vdots \\ \sum_{t=1}^{T}\delta_{St}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}.$$

$\mathbf{U}'_{\xi r}$ is a $(S-1)\times S$ matrix with rank $S-1$ and the previous homogenous linear system has an infinite number of solutions. However, as the $\sum_{i=1}^{S}\sum_{t=1}^{T}\delta_{it} = 0$ sum-to-zero constraint is satisfied, adding this constraint to the linear system:

$$\begin{bmatrix}\mathbf{U}'_{\xi r}\\ \mathbf{1}'_S\end{bmatrix}\begin{pmatrix}\sum_{t=1}^{T}\delta_{1t}\\ \vdots \\ \sum_{t=1}^{T}\delta_{St}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix} \Longleftrightarrow \begin{pmatrix}\sum_{t=1}^{T}\delta_{1t}\\ \vdots \\ \sum_{t=1}^{T}\delta_{St}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}$$
$$\Longleftrightarrow \sum_{t=1}^{T}\delta_{it} = 0,\ \forall i.$$

- In a similar way, it can be shown that

$$(\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)'\boldsymbol{\delta} = \mathbf{U}'_{\gamma r}\begin{pmatrix}\mathbf{1}'_S & \cdots & 0\\ 0 & \ddots & 0\\ 0 & 0 & \mathbf{1}'_S\end{pmatrix}$$

$$\boldsymbol{\delta} = \mathbf{U}'_{\gamma r}(\mathbf{I}_\gamma \otimes \mathbf{1}'_S)\boldsymbol{\delta} = \mathbf{U}'_{\gamma r}\begin{pmatrix}\sum_{i=1}^{S}\delta_{i1}\\ \vdots \\ \sum_{i=1}^{S}\delta_{iT}\end{pmatrix},$$

and consequently,

$$(\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)'\boldsymbol{\delta} = \mathbf{0} \Longleftrightarrow \mathbf{U}'_{\gamma r}\begin{pmatrix}\sum_{i=1}^{S}\delta_{i1}\\ \vdots \\ \sum_{i=1}^{S}\delta_{iT}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}.$$

$\mathbf{U}'_{\gamma r}$ is a $(T-1)\times T$ matrix with rank $T-1$, and the previous homogenous linear system has an infinite number of solutions. However, as the $\sum_{i=1}^{S}\sum_{t=1}^{T}\delta_{it} = 0$ sum-to-zero constraint is satisfied, adding this constraint to the linear system

$$\begin{bmatrix}\mathbf{U}'_{\gamma r}\\ \mathbf{1}'_T\end{bmatrix}\begin{pmatrix}\sum_{i=1}^{S}\delta_{i1}\\ \vdots \\ \sum_{i=1}^{S}\delta_{iT}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix} \Longleftrightarrow \begin{pmatrix}\sum_{i=1}^{S}\delta_{i1}\\ \vdots \\ \sum_{i=1}^{S}\delta_{iT}\end{pmatrix} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}$$
$$\Longleftrightarrow \sum_{i=1}^{S}\delta_{it} = 0,\ \forall t.$$

Appendix 1.2

Given the spatio-temporal model of Eq. (8)

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi} + (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} + \mathbf{I}_\delta\boldsymbol{\delta},$$

where η is the log of the global risk, the spatial random effect $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma^2_\xi \mathbf{Q}^-_\xi\right)$ follows an ICAR distribution, the temporal random effect $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \sigma^2_\gamma \mathbf{Q}^-_\gamma\right)$ follows a RW2 distribution, and the interaction random effects $\boldsymbol{\delta} \sim N\left(\mathbf{0}, \sigma^2_\delta (\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)^-\right)$ are completely structured (Type IV interaction) in space and time, the linear constraints that make this model identifiable are

$$\sum_{i=0}^{S} \xi_i = 0,\ \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \begin{array}{l} \sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for } i = 1, \ldots, S \\ \sum_{i=1}^{S} \delta_{it} = 0, \quad \text{for } t = 1, \ldots, T \end{array}$$

Additionally, the constraints $\sum_{t=1}^{T}\sum_{i=1}^{S} t\delta_{it} = 0$ have to be considered, but they are automatically placed with $\sum_{i=1}^{S} \delta_{it} = 0$, as $\sum_{t=1}^{T}\sum_{i=1}^{S} t\delta_{it} = \sum_{t=1}^{T} t \sum_{i=1}^{S} \delta_{it} = 0$.

If the temporal random effect $\boldsymbol{\gamma}$ follows a RW2 distribution, the random effects of Eq. (8) can be reparameterized using the spectral decomposition of the covariance matrix, so this model becomes

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + [\mathbf{X}_\xi : \mathbf{Z}_\xi]\begin{bmatrix} \beta_\xi \\ \boldsymbol{\alpha}_\xi \end{bmatrix} + [\mathbf{X}_\gamma : \mathbf{Z}_\gamma]\begin{bmatrix} \beta_\gamma \\ \boldsymbol{\alpha}_\gamma \end{bmatrix} + [\mathbf{X}_\delta : \mathbf{Z}_\delta]\begin{bmatrix} \boldsymbol{\beta}_\delta \\ \boldsymbol{\alpha}_\delta \end{bmatrix},$$

where

$$\begin{aligned}
&\mathbf{X}_\xi = (\mathbf{1}_T \otimes \mathbf{I}_\xi)\mathbf{U}_{\xi n} = \mathbf{1}_{TS}, \quad \beta_\xi = \mathbf{U}'_{\xi n}\boldsymbol{\xi} = \mathbf{1}'_S\boldsymbol{\xi},\\
&\mathbf{Z}_\xi = (\mathbf{1}_T \otimes \mathbf{I}_\xi)\mathbf{U}_{\xi r} = \mathbf{1}_T \otimes \mathbf{U}_{\xi r}, \quad \boldsymbol{\alpha}_\xi = \mathbf{U}'_{\xi r}\boldsymbol{\xi},\\
&\mathbf{X}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma n} = [\mathbf{1}_{TS} : \mathbf{t}^* \otimes \mathbf{1}_S], \quad \boldsymbol{\beta}_\gamma = \mathbf{U}'_{\gamma n}\boldsymbol{\gamma} = [\mathbf{1}_T : \mathbf{t}^*]'\boldsymbol{\gamma},\\
&\mathbf{Z}_\gamma = (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\mathbf{U}_{\gamma r} = \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S, \quad \boldsymbol{\alpha}_\gamma = \mathbf{U}'_{\gamma r}\boldsymbol{\gamma},\\
&\mathbf{X}_\delta = \mathbf{U}_{\delta n} = [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{U}_{\xi r}],\\
&\boldsymbol{\beta}_\delta = \mathbf{U}'_{\delta n}\boldsymbol{\delta}, \mathbf{Z}_\delta = \mathbf{U}_{\delta r} = [\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r}], \quad \boldsymbol{\alpha}_\delta = \mathbf{U}'_{\delta r}\boldsymbol{\delta}.
\end{aligned}$$

Consequently, this model can be re-expressed as

$$\begin{aligned}\log \mathbf{r} = {} & \mathbf{1}_{TS}\eta + \mathbf{1}_{TS}\beta_\xi + (\mathbf{1}_T \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\xi + [\mathbf{1}_{TS} : \mathbf{t}^* \otimes \mathbf{1}_S]\boldsymbol{\beta}_\gamma \\ & + (\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)\boldsymbol{\alpha}_\gamma + [\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S : \mathbf{t}^* \\ & \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{U}_{\xi r}]\boldsymbol{\beta}_\delta + (\mathbf{U}_{\gamma r} \otimes \mathbf{U}_{\xi r})\boldsymbol{\alpha}_\delta.\end{aligned}$$

where $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma^2_\xi \widetilde{\boldsymbol{\Sigma}}^{-1}_\xi\right)$, $\boldsymbol{\alpha}_\gamma \sim N\left(\mathbf{0}, \sigma^2_\gamma \widetilde{\boldsymbol{\Sigma}}^{-1}_\gamma\right)$ and $\boldsymbol{\alpha}_\delta \sim N\left(\mathbf{0}, \sigma^2_\delta \widetilde{\boldsymbol{\Sigma}}^{-1}_\delta\right)$.

To obtain the identifiable model of Eq. (12), the repeated columns $\mathbf{1}_{TS}$ (corresponding to $\boldsymbol{\beta}_\gamma$), $\mathbf{1}_T \otimes \mathbf{U}_{\xi r}$, $\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S$ and $\mathbf{t}^* \otimes \mathbf{1}_S$ (corresponding to $\boldsymbol{\beta}_\delta$) must be removed, which is equivalent to set $\mathbf{1}'_S\boldsymbol{\xi} = 0$, $\mathbf{1}'_T\boldsymbol{\gamma} = 0$ and $[\mathbf{1}_{TS} : \mathbf{1}_T \otimes \mathbf{U}_{\xi r} : \mathbf{U}_{\gamma r} \otimes \mathbf{1}_S : \mathbf{t}^* \otimes \mathbf{1}_S]'\boldsymbol{\delta} = \mathbf{0}$.

Similar to the RW1 case, it can be shown that:

- $\mathbf{1}'_S\boldsymbol{\xi} = 0 \iff \sum_{i=1}^{S} \xi_i = 0$, $\quad \mathbf{1}'_T\boldsymbol{\gamma} = 0 \iff \sum_{t=1}^{T} \gamma_t = 0$ and $\mathbf{1}'_{TS}\boldsymbol{\delta} = 0 \iff \sum_{i=1}^{S}\sum_{t=1}^{T} \delta_{it} = 0$.
- $(\mathbf{t}^* \otimes \mathbf{1}_S)'\boldsymbol{\delta} = 0 \iff (\mathbf{1}'_S, 2\mathbf{1}'_S, \ldots, T\mathbf{1}'_S)\boldsymbol{\delta} = 0 \iff \sum_{i=1}^{S}\sum_{t=1}^{T} t\delta_{it} = 0.$
- $(\mathbf{1}_T \otimes \mathbf{U}_{\xi r})'\boldsymbol{\delta} = \mathbf{0} \iff \mathbf{U}'_{\xi r}\begin{pmatrix} \sum_{t=1}^{T} \delta_{1t} \\ \vdots \\ \sum_{t=1}^{T} \delta_{St} \end{pmatrix} = \mathbf{0} \iff \sum_{t=1}^{T} \delta_{it} = 0, \forall i.$
- Finally,

$$(\mathbf{U}_{\gamma r} \otimes \mathbf{1}_S)'\boldsymbol{\delta} = \mathbf{0} \iff \mathbf{U}'_{\gamma r}\begin{pmatrix} \sum_{i=1}^{S} \delta_{1i} \\ \vdots \\ \sum_{i=1}^{S} \delta_{Ti} \end{pmatrix} = \mathbf{0},$$

$\mathbf{U}'_{\gamma r}$ is a $(T-2) \times T$ matrix with rank $T-2$, and the previous homogenous linear system has an infinite number of solutions. However, adding both $\sum_{i=1}^{S}\sum_{t=1}^{T} \delta_{it} = 0$ and $\sum_{i=1}^{S}\sum_{t=1}^{T} t\delta_{it} = 0$ sum-to-zero constraints to the linear system

$$\begin{bmatrix} \mathbf{U}'_{\gamma r} \\ \mathbf{1}'_T \\ \mathbf{t}^{*\prime} \end{bmatrix}\begin{pmatrix} \sum_{i=1}^{S} \delta_{i1} \\ \vdots \\ \sum_{i=1}^{S} \delta_{iT} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \iff \begin{pmatrix} \sum_{i=1}^{S} \delta_{i1} \\ \vdots \\ \sum_{i=1}^{S} \delta_{iT} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\iff \sum_{i=1}^{S} \delta_{it} = 0, \forall t.$$

The reader should note that if the Leroux et al. (1999) reparameterization is used, the same sum-to-zero constraints have to be considered because the linear combination $\sum_{i=1}^{S} \xi_i$ is in the span of the intercept.

Appendix 2

In this section we provide the constraints for interactions of Type I, II, and III.

Appendix 2.1: RW1

Given the spatio-temporal model of Eq. (8)

$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi} + (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} + \mathbf{I}_\delta\boldsymbol{\delta},$$

where η is the log of the global risk, the spatial random effect $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma^2_\xi \mathbf{Q}^-_\xi\right)$ follows an ICAR distribution, the temporal random effect $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \sigma^2_\gamma \mathbf{Q}^-_\gamma\right)$ follows a RW1 distribution.

1. If the interaction effects are unstructured in space and time, that is, the interaction is of Type I with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{I}_\delta)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \sum_{t=1}^{T}\sum_{i=1}^{S} \delta_{it} = 0$$
2. If the interaction effects are unstructured in space and structured in time, that is, the interaction is of Type II with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 (\mathbf{Q}_\gamma \otimes \mathbf{I}_\xi)^-)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \sum_{t=1}^{T} \delta_{it} = 0, \quad i = 1, \ldots, S.$$
3. If the interaction effects are structured in space and unstructured in time, that is, the interaction is of Type III with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 (\mathbf{I}_\gamma \otimes \mathbf{Q}_\xi)^-)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \sum_{i=1}^{S} \delta_{it} = 0, \quad t = 1, \ldots, T.$$

Appendix 2.2: RW2

Given the spatio-temporal model of Eq. (8)
$$\log \mathbf{r} = \mathbf{1}_{TS}\eta + (\mathbf{1}_T \otimes \mathbf{I}_\xi)\boldsymbol{\xi} + (\mathbf{I}_\gamma \otimes \mathbf{1}_S)\boldsymbol{\gamma} + \mathbf{I}_\delta \boldsymbol{\delta},$$
where η is the log of the global risk, the spatial random effect $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma_\xi^2 \mathbf{Q}_\xi^-\right)$ follows an ICAR distribution, the temporal random effect $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \sigma_\gamma^2 \mathbf{Q}_\gamma^-\right)$ follows a RW2 distribution.

1. If the interaction effects are unstructured in space and time, that is, the interaction is of Type I with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{I}_\delta)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and} \quad \begin{array}{l} \sum_{t=1}^{T}\sum_{i=1}^{S} \delta_{it} = 0, \\ \sum_{t=1}^{T}\sum_{i=1}^{S} t\delta_{it} = 0. \end{array}$$
2. If the interaction effects are unstructured in space and structured in time, that is, the interaction is of Type II with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 (\mathbf{Q}_\gamma \otimes \mathbf{I}_\xi)^-)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0, \quad \sum_{t=1}^{T} t\gamma_t = 0 \quad \text{and}$$
$$\sum_{t=1}^{T} \delta_{it} = 0, \quad i = 1, \ldots, S$$
3. If the interaction effects are structured in space and unstructured in time, that is, the interaction is of Type III with $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 (\mathbf{I}_\gamma \otimes \mathbf{Q}_\xi)^-)$, the linear constraints that make this model identifiable are
$$\sum_{i=1}^{S} \xi_i = 0, \quad \sum_{t=1}^{T} \gamma_t = 0 \quad \text{and}$$
$$\sum_{i=1}^{S} \delta_{it} = 0, \quad \text{for} \quad t = 1, \ldots, T.$$

Appendix 3

This section shows that the variance of the intercept is inflated if the LCAR prior is used for the spatial random effect and the model is fitted using PQL with appropriate constraints. If the model is reparameterized, the variance is not inflated. It also shows that the ICAR prior does not present this problem. We consider this spatial linear mixed model:
$$\mathbf{Y} = \mathbf{1}_S \eta + \boldsymbol{\xi} + \boldsymbol{\epsilon}. \tag{13}$$

Appendix 3.1: ICAR prior

Assume an ICAR prior for the spatial random effect $\boldsymbol{\xi}$, that is, $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma_\xi^2 \mathbf{Q}_\xi^-\right)$, and $\boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right)$. Then
$$\mathrm{var}(Y) = \mathbf{V} = \sigma_\epsilon^2 \mathbf{I}_S + \sigma_\xi^2 \mathbf{Q}_\xi^- = \sigma_\epsilon^2 (\mathbf{I}_S + k\mathbf{U}_\xi \boldsymbol{\Sigma}_\xi^- \mathbf{U}_\xi'),$$
where $\mathbf{I}_S$ is an $S \times S$ identity matrix, $\mathbf{U}_\xi = [\mathbf{U}_{\xi n} : \mathbf{U}_{\xi r}] = [\mathbf{1}_S/\sqrt{S} : \mathbf{U}_{\xi r}]$, and $k = \sigma_\xi^2/\sigma_\epsilon^2$. Clearly
$$\mathbf{V}^{-1} = \frac{1}{\sigma_\epsilon^2}(\mathbf{I}_S + k\mathbf{U}_\xi \boldsymbol{\Sigma}_\xi^- \mathbf{U}_\xi')^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{U}_\xi (\mathbf{I}_S + k\boldsymbol{\Sigma}_\xi^-)^{-1}\mathbf{U}_\xi'.$$
Then
$$\begin{aligned}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) &= \frac{1}{\sigma_\epsilon^2}\mathbf{1}_S'\mathbf{U}_\xi(\mathbf{I}_S + k\boldsymbol{\Sigma}_\xi^-)^{-1}\mathbf{U}_\xi'\mathbf{1}_S \\ &= \frac{1}{\sigma_\epsilon^2}\left[S/\sqrt{S}, 0, \ldots, 0\right] \mathrm{diag} \\ &\quad (1, 1 + k/d_2, \ldots, 1 + k/d_S)^{-1}\left[S/\sqrt{S}, 0, \ldots, 0\right]' = \frac{S}{\sigma_\epsilon^2},\end{aligned}$$
where $\mathbf{X} = 1_S$, $d_2, \ldots, d_S$ are the non-null eigenvalues of $\mathbf{Q}_\xi$ (note that $d_1 = 0$), and finally
$$\mathrm{var}(\hat{\eta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \frac{\sigma_\epsilon^2}{S}.$$

Now reparameterize Model (13). Then
$\mathbf{Y} = \mathbf{1}_S\eta + \mathbf{U}_{\xi r}\boldsymbol{\alpha}_\xi + \boldsymbol{\epsilon}$,

where $\boldsymbol{\alpha} \sim N\left(\mathbf{0}, \sigma^2_\xi \widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\right)$. Then,

$$\mathrm{var}(Y) = \mathbf{V} = \sigma^2_\epsilon\left(\mathbf{I}_S + k\mathbf{U}_{\xi r}\widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\mathbf{U}'_{\xi r}\right).$$

Using matrix inversion formulas and noting that $\mathbf{U}'_{\xi r}\mathbf{U}_{\xi r} = \mathbf{I}_{S-1}$, it follows that

$$\begin{aligned}\mathbf{V}^{-1} &= \left[\sigma^2_\epsilon(\mathbf{I}_S + k\mathbf{U}_{\xi r}\widetilde{\boldsymbol{\Sigma}}_\xi^{-1}\mathbf{U}'_{\xi r})\right]^{-1}\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I}_S - \mathbf{I}_S\mathbf{U}_{\xi r}(\mathbf{U}'_{\xi r}\mathbf{I}_S\mathbf{U}_{\xi r} + k^{-1}\widetilde{\boldsymbol{\Sigma}}_\xi)^{-1}\mathbf{U}'_{\xi r}\mathbf{I}_S\right]\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I}_S - \mathbf{U}_{\xi r}(\mathbf{I}_{S-1} + k^{-1}\widetilde{\boldsymbol{\Sigma}}_\xi)^{-1}\mathbf{U}'_{\xi r}\right],\end{aligned}$$

and

$$\begin{aligned}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right) &= \frac{1}{\sigma^2_\epsilon}\mathbf{1}'_S\left[\mathbf{I}_S - \mathbf{U}_{\xi r}(\mathbf{I}_{S-1} + k^{-1}\widetilde{\boldsymbol{\Sigma}}_\xi)^{-1}\mathbf{U}'_{\xi r}\right]\mathbf{1}_S\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{1}'_S\mathbf{1}_S - \mathbf{1}'_S\mathbf{U}_{\xi r}(\mathbf{I}_{S-1} + k^{-1}\widetilde{\boldsymbol{\Sigma}}_\xi)^{-1}\mathbf{U}'_{\xi r}\mathbf{1}_S\right].\\ &= \frac{1}{\sigma^2_\epsilon}(S-0) = \frac{S}{\sigma^2_\epsilon},\end{aligned}$$

because $\mathbf{1}_S$ and the columns of $\mathbf{U}_{\xi r}$ are orthogonal. Consequently,

$$\mathrm{var}(\hat{\eta}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} = \frac{\sigma^2_\epsilon}{S},$$

and the variance is the same if the model is fitted with appropriate sum-to-zero constraints or in the reparameterized version.

Appendix 3.2: LCAR prior

Assume an LCAR prior for the spatial random effect $\boldsymbol{\xi}$, that is $\boldsymbol{\xi} \sim N\left(\mathbf{0}, \sigma^2_\xi\mathbf{D}_\xi^{-1}\right)$, $\mathbf{D} = \lambda_\xi\mathbf{Q}_\xi + (1-\lambda_\xi)\mathbf{I}_S$. Then

$$\begin{aligned}\mathrm{var}(Y) = \mathbf{V} &= \sigma^2_\epsilon\mathbf{I}_S + \sigma^2_\xi\mathbf{D}_\xi^{-1} = \sigma^2_\epsilon(\mathbf{I}_S + k\mathbf{D}^{-1})\\ &= \sigma^2_\epsilon\left[\mathbf{I}_S + k\mathbf{U}_\xi((1-\lambda_\xi)\mathbf{I}_S + \lambda_\xi\boldsymbol{\Sigma}_\xi)^{-1}\mathbf{U}'_\xi\right],\end{aligned}$$

where as before, $\mathbf{I}_S$ is an $S \times S$ identity matrix, $\mathbf{U}_\xi = [\mathbf{U}_{\xi n} : \mathbf{U}_{\xi r}] = [\mathbf{1}_S/\sqrt{S} : \mathbf{U}_{\xi r}]$, and $k = \sigma^2_\xi/\sigma^2_\epsilon$. Clearly

$$\begin{aligned}\mathbf{V}^{-1} &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I}_S + k\mathbf{U}_\xi((1-\lambda_\xi)\mathbf{I}_S + \lambda_\xi\boldsymbol{\Sigma}_\xi)^{-1}\mathbf{U}'_\xi\right]^{-1}\\ &= \frac{1}{\sigma^2_\epsilon}\mathbf{U}_\xi\left[\mathbf{I}_S + k((1-\lambda_\xi)\mathbf{I}_S + \lambda_\xi\boldsymbol{\Sigma}_\xi)^{-1}\right]^{-1}\mathbf{U}'_\xi\\ &= \frac{1}{\sigma^2_\epsilon}\mathbf{U}_\xi\ \mathrm{diag}\left(1 + \frac{k}{\lambda_\xi d_1 + (1-\lambda_\xi)}, \ldots, 1 + \frac{k}{\lambda_\xi d_S + (1-\lambda_\xi)}\right)^{-1}\mathbf{U}'_\xi.\end{aligned}$$

Note that the eigenvalues of the precision matrix $\lambda_\xi\mathbf{Q}_\xi + (1-\lambda_\xi\mathbf{I}_S)$ are all positive whenever $\lambda_\xi < 1$. They take the form $\lambda_\xi d_i + (1-\lambda_\xi)$, $i = 1, \ldots, S$, where $d_1 = 0$ and $d_i > 0$, $i = 2, \ldots, S$ are the eigenvalues of $\mathbf{Q}_\xi$. Then

$$\begin{aligned}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) &= \frac{1}{\sigma^2_\epsilon}\mathbf{1}'_S\mathbf{U}_\xi\ \mathrm{diag}\left(1 + \frac{k}{(1-\lambda_\xi)}, \ldots, 1 + \frac{k}{\lambda_\xi d_S + (1-\lambda_\xi)}\right)^{-1}\mathbf{1}_S = \frac{1}{\sigma^2_\epsilon}\left[S/\sqrt{S}, 0, \ldots, 0\right]\\ &\quad \mathrm{diag}\left(1 + \frac{k}{(1-\lambda_\xi)}, \ldots, 1 + \frac{k}{\lambda_\xi d_S + (1-\lambda_\xi)}\right)^{-1}\\ &\quad \times\left[S/\sqrt{S}, 0, \ldots, 0\right]' = \frac{S}{\sigma^2_\epsilon}\left(1 + \frac{k}{(1-\lambda_\xi)}\right)^{-1}.\end{aligned}$$

Finally

$$\mathrm{var}(\hat{\eta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \frac{\sigma^2_\epsilon}{S}\left(1 + \frac{k}{(1-\lambda_\xi)}\right).$$

Now reparameterize Model (13). Then

$$\mathbf{Y} = \mathbf{1}_S\eta + \mathbf{U}_{\xi r}\boldsymbol{\alpha}_\xi + \boldsymbol{\epsilon},$$

where $\boldsymbol{\alpha}_\xi \sim N\left(\mathbf{0}, \sigma^2_\xi(\lambda_\xi\widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{S-1})^{-1}\right)$. Denoting by $\widetilde{\mathbf{D}} = \lambda_\xi\widetilde{\boldsymbol{\Sigma}}_\xi + (1-\lambda_\xi)\mathbf{I}_{S-1}$,

$$\mathrm{var}(Y) = \mathbf{V} = \sigma^2_\epsilon(\mathbf{I}_S + k\mathbf{U}_{\xi r}\widetilde{\mathbf{D}}^{-1}\mathbf{U}'_{\xi r}).$$

Using matrix inversion formulas and taking into account that $\mathbf{U}'_{\xi r}\mathbf{U}_{\xi r} = \mathbf{I}_{S-1}$,

$$\begin{aligned}\mathbf{V}^{-1} &= \left[\sigma^2_\epsilon(\mathbf{I}_S + k\mathbf{U}_{\xi r}\widetilde{\mathbf{D}}^{-1}\mathbf{U}'_{\xi r})\right]^{-1}\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I}_S - \mathbf{I}_S\mathbf{U}_{\xi r}\left(\mathbf{U}'_{\xi r}\mathbf{I}_S\mathbf{U}_{\xi r} + k^{-1}\widetilde{\mathbf{D}}\right)^{-1}\mathbf{U}'_{\xi r}\mathbf{I}_S\right]\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I}_S - \mathbf{U}_{\xi r}\left(\mathbf{I}_{S-1} + k^{-1}\widetilde{\mathbf{D}}\right)^{-1}\mathbf{U}'_{\xi r}\right],\end{aligned}$$

and

$$\begin{aligned}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) &= \frac{1}{\sigma^2_\epsilon}\mathbf{1}'_S\left[\mathbf{I}_S - \mathbf{U}_{\xi r}\left(\mathbf{I}_{S-1} + k^{-1}\widetilde{\mathbf{D}}\right)^{-1}\mathbf{U}'_{\xi r}\right]\mathbf{1}_S\\ &= \frac{1}{\sigma^2_\epsilon}\left[\mathbf{1}'_S\mathbf{1}_S - \mathbf{1}'_S\mathbf{U}_{\xi r}\left(\mathbf{I}_{S-1} + k^{-1}\widetilde{\mathbf{D}}\right)^{-1}\mathbf{U}'_{\xi r}\mathbf{1}_S\right].\\ &= \frac{1}{\sigma^2_\epsilon}(S-0) = \frac{S}{\sigma^2_\epsilon}.\end{aligned}$$

Consequently,

$$\mathrm{var}(\hat{\eta}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} = \frac{\sigma^2_\epsilon}{S}.$$

Thus it is clear that if the model is fitted with sum-to-zero constraints the variance of the intercept estimator is inflated. This is avoided by reparameterizing the model.

Appendix 4

The `R` code to fit the spatio-temporal Model (4) in INLA considering different types of interactions is detailed below. First, the data frame (or list) containing the variables of the model has to be defined

```
> Data <- data.frame(O=<observed>, E=<expected>,
+                    ID.area=rep(1:S,T),
+                    ID.year=rep(1:T,each=S),
+                    ID.area.year=seq(1,T*S))
```

where `observed` and `expected` are the vectors of observed and expected deaths (or incident cases) respectively, and `S` and `T` are the number of small areas and time periods for which data are available (`S=50` provinces and `T=21` years for breast cancer mortality data). Note that the data must be ordered according to the Kronecker product given for the structure matrix of the space-time interaction random effect $\boldsymbol{\delta}$. For details about how to introduce the data or how the IDs must be specified in INLA, see the examples and tutorials in http://www.r-inla.org/examples.

Then, we define the spatial neighborhood matrix $\mathbf{Q}_{\xi}$ and the necessary structure matrix to implement the LCAR prior using the `''generic1''` model (see Ugarte et al. 2014) as

```
> g <- inla.read.graph("prov_nb.inla")
> Q.xi <- matrix(0, g$n, g$n)
> for (i in 1:g$n){
+   Q.xi[i,i]=g$nnbs[[i]]
+   Q.xi[i,g$nbs[[i]]]=-1
+ }
>
> Q.Leroux <- diag(S)-Q.xi
```

where `''prov_nb.inla''` is an `inla.graph` object containing the neighbouring structure of the Spanish provinces. Similarly, we define the temporal structure matrix $\mathbf{Q}_{\gamma}$ of a random walk of first or second order as

```
> D1 <- diff(diag(T),differences=1)
> Q.gammaRW1 <- t(D1)%*%D1
>
> D2 <- diff(diag(T),differences=2)
> Q.gammaRW2 <- t(D2)%*%D2
```

The `''expression''` statement of `R-INLA` can be used to implement uniform distributions over the standard deviations of the random effects as

```
> sdunif = "expression:
+  logdens = -log_precision/2;
+  return(logdens)"
```

or standard uniform distribution for the spatial smoothing parameter λ_{ξ}

```
> lunif = "expression:
+  a = 1;
+  b = 1;
+  beta = exp(theta)/(1+exp(theta));
+  logdens = lgamma(a+b)-lgamma(a)-lgamma(b)+(a-1)*beta+(b-1)*(1-beta);
+  log_jacobian = log(beta*(1-beta));
+  return(logdens+log_jacobian)"
```

Details about how these priors are derived and their implementation in `R-INLA` can be found in Ugarte et al. (2016). See the web page http://www.r-inla.org/models/priors for descriptions and examples of available prior distributions already implemented in `R-INLA`.

To define the `formula` object of each model, the `f()` function is used. For the spatially structured random effect the `''generic1''` model can be used to implement the LCAR prior distribution, while `''rw1''` and `''rw2''` models are available to define first and second order random walk priors for the temporally structured random effect. For the spatio-temporal interaction effect the `''generic0''` model can be used, which defines a generic Gaussian prior distribution in terms of a structure matrix given by the user.

The formula for a model with **Type IV interaction and RW1 prior for time** is defined as

```
> R <- kronecker(Q.gammaRW1,Q.xi)
> r.def <- S+T-1
> A1 <- kronecker(matrix(1,1,T),diag(S))
> A2 <- kronecker(diag(T),matrix(1,1,S))
> A.constr <- rbind(A1,A2)
```

```
> formula <- 0 ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw1", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif))) +
+                f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=A.constr, e=rep(0,S+T)))
```

where `R` is the structure matrix $\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi$, whose rank deficiency is specified in the **rankdef** argument. The linear constraints that make this model identifiable (see "Appendix 1") are given by the **constr=TRUE** (a sum-to-zero constraint over the random effect) and the `extraconstr` arguments. Note that

$$\sum_{t=1}^{T} \delta_{it} = 0, \quad \text{for} \quad i = 1,\ldots,S \Longleftrightarrow (\mathbf{1}'_T \otimes \mathbf{I}_\xi)\boldsymbol{\delta} = 0,$$

and

$$\sum_{i=1}^{S} \delta_{it} = 0, \quad \text{for} \quad t = 1,\ldots,T \Longleftrightarrow (\mathbf{I}_\gamma \otimes \mathbf{1}'_S)\boldsymbol{\delta} = 0.$$

Similarly, the formula for a model with **Type IV interaction and RW2 prior for time** is defined as

```
> R <- kronecker(Q.gammaRW2,Q.xi)
> r.def <- 2*S+T-2
> A1 <- kronecker(matrix(1,1,T),diag(S))
> A2 <- kronecker(diag(T),matrix(1,1,S))
> A.constr <- rbind(A1,A2)
```

```
> formula <- 0 ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw2", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif))) +
+                f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=A.constr, e=rep(0,S+T)))
```

Finally, we run the INLA algorithm with a call to the `inla()` function as

```
> inla(formula, family="poisson", data=Data, E=E,
+      control.predictor=list(compute=TRUE),
+      control.compute=list(dic=TRUE),
+      control.inla=list(strategy=<strategy>))
```

where the approximation strategy to compute the marginal posteriors of the latent Gaussian field is specified with the **strategy** argument. Three different approaches are possible (see Rue et al. 2009): a Gaussian approximation (`''gaussian''`), a simplified Laplace approximation (`''simplified.laplace''`, default option in INLA) and a full Laplace approximation (`''laplace''`).

The `formula` object for the other types of interactions are detailed below.

Type I interaction and RW1 prior for time

```
> formula <- O ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw1", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif))) +
+                f(ID.area.year, model="iid", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif)))
```

Type I interaction and RW2 prior for time

```
> formula <- O ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw2", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif))) +
+                f(ID.area.year, model="iid", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=matrix(rep(1:T,S),1,S*T),e=0))
```

Type II interaction and RW1 prior for time

```
> R <- kronecker(Q.gammaRW1,diag(S))
> r.def <- S
> A.constr <- kronecker(matrix(1,1,T),diag(S))
```

```
> formula <- O ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw1", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif))) +
+                f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=A.constr, e=rep(0,S)))
```

Type II interaction and RW2 prior for time

```
> R <- kronecker(Q.gammaRW2,diag(S))
> r.def <- 2*S
> A.constr <- kronecker(matrix(1,1,T),diag(S))
```

```
> formula <- O ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                f(ID.year, model="rw2", constr=TRUE,
+                  hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=matrix(1:T,1,T),e=0)) +
+                f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                  constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                  extraconstr=list(A=A.constr, e=rep(0,S)))
```

Type III interaction and RW1 prior for time

```
> R <- kronecker(diag(T),Q.xi)
> r.def <- T
> A.constr <- kronecker(diag(T),matrix(1,1,S))
```

```
> formula <- 0 ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                   hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                 f(ID.year, model="rw1", constr=TRUE,
+                   hyper=list(prec=list(prior=sdunif))) +
+                 f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                   constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                   extraconstr=list(A=A.constr, e=rep(0,T)))
```

Type III interaction and RW2 prior for time

```
> R <- kronecker(diag(T),Q.xi)
> r.def <- T
> A.constr <- kronecker(diag(T),matrix(1,1,S))
```

```
> formula <- 0 ~ f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
+                   hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
+                 f(ID.year, model="rw2", constr=TRUE,
+                   hyper=list(prec=list(prior=sdunif)),
+                 f(ID.area.year, model="generic0", Cmatrix=R, rankdef=r.def,
+                   constr=TRUE, hyper=list(prec=list(prior=sdunif)),
+                   extraconstr=list(A=A.constr, e=rep(0,T)))
```

References

Adin A, Martínez-Beneito MA, Botella-Rocamora P, Goicoa T, Ugarte MD (2016) Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. Stoch Environ Res Risk Assess. 31(2):403–415. doi:10.1007/s00477-016-1269-8

Ainsworth L, Dean C (2006) Approximate inference for disease mapping. Comput Stat Data Anal 50(10):2552–2570

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M (1995) Bayesian analysis of space time variation in disease risk. Stat Med 14(21–22):2433–2443

Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43(1):1–20

Best N, Richardson S, Thomson A (2005) A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res 14(1):35–59

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88(421):9–25

Dean C, Ugarte MD, Militino AF (2004) Penalized quasi-likelihood with spatially correlated data. Comput Stat Data Anal 45(2):235–248

Eberly LE, Carlin BP et al (2000) Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat Med 19(1718):2279–2294

Etxeberria J, Goicoa T, Ugarte MD, Militino AF (2014) Evaluating space-time models for short-term cancer mortality risk predictions in small areas. Biom J 56(3):383–402

Gelfand AE, Sahu SK (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. J Am Stat Asso 94(445):247–253

Gilks W (2005) Markov chain Monte Carlo. Wiley, Hoboken

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. J Am Stat Assoc 72(358):320–338

Harville DA (2008) Matrix algebra from a statistician's perspective, 2nd edn. Springer, New York

Hodges JS, Reich BJ (2010) Adding spatially-correlated errors can mess up the fixed effect you love. Am Stat 64(4):325–334

Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. Stat Med 19(17–18):2555–2567

Knorr-Held L, Besag J (1998) Modelling risk from a disease in time and space. Stat Med 17(18):2045–2060

Knorr-Held L, Rue H (2002) On block updating in Markov random field models for disease mapping. Scand J Stat 29(4):597–614

Leroux BG, Lei X, Breslow N (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran M, Berry D (eds) Statistical models in epidemiology, the environment, and clinical trials. Springer, New York, pp 179–192

MacNab YC (2007) Spline smoothing in Bayesian disease mapping. Environmetrics 18(7):727–744

MacNab YC, Dean C (2001) Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. Biometrics 57(3):949–956

MacNab YC, Gustafson P (2007) Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. Stat Med 26(24):4455–4474

Martínez-Beneito M, López-Quilez A, Botella-Rocamora P (2008) An autoregressive approach to spatio-temporal disease mapping. Stat Med 27(15):2874–2889

Martino S, Rue H (2009) Implementing approximate Bayesian inference using integrated nested laplace approximation: a manual for the inla program. Department of Mathematical Sciences, NTNU, Norway

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/

Reich BJ, Hodges JS, Zadnik V (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 62(4):1197–1206

Rue H, Held L (2005) Gaussian Markov random fields: theory and applications, vol 104. CRC press, Boca Raton

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B (Stat Methodol) 71(2):319–392

Schmid V, Held L (2004) Bayesian extrapolation of space-time trends in cancer registry data. Biometrics 60(4):1034–1042

Schrödle B, Held L (2011) Spatio-temporal disease mapping using INLA. Environmetrics 22(6):725–734

Schrödle B, Held L, Riebler A, Danuser J (2011) Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. J R Stat Soc Ser C (Appl Stat) 60(2):261–279

Ugarte M, Militino A, Goicoa T (2008) Prediction error estimators in empirical Bayes disease mapping. Environmetrics 19(3):287–300

Ugarte M, Goicoa T, Ibáñez B, Militino A (2009a) Evaluating the performance of spatio-temporal Bayesian models in disease mapping. Environmetrics 20(6):647–665

Ugarte MD, Goicoa T, Militino AF (2009b) Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. Comput Stat Data Anal 53(8):2938–2949

Ugarte MD, Goicoa T, Militino AF (2010) Spatio-temporal modeling of mortality risks using penalized splines. Environmetrics 21(3–4):270–289

Ugarte M, Goicoa T, Etxeberria J, Militino A (2012) A p-spline anova type model in space-time disease mapping. Stoch Environ Res Risk Assess 26(6):835–845

Ugarte MD, Adin A, Goicoa T, Militino AF (2014) On fitting spatio-temporal disease mapping models using approximate Bayesian inference. Stat Methods Med Res 23(6):507–530. doi:10.1177/0962280214527528

Ugarte MD, Adin A, Goicoa T (2016) Two-level spatially structured models in spatio-temporal disease mapping. Stat Methods Med Res 25(4):1080–1100