Space-time interactions in Bayesian disease mapping with recent tools: Making things easier for practitioners

Arantxa Urdangarin[1,2], Tomás Goicoa[1,2,3], María Dolores Ugarte[1,2,3*]

[1] *Department of Statistics, Computer Science, and Mathematics, Public University of Navarre, Spain.*

[2] *INAMAT[2] (Institute for Advanced Materials and Mathematics) , Public University of Navarre, Spain.*

[3] *Institute of Health Research, IdisNA, Spain.*

∗**Corresponding author:** María Dolores Ugarte, Department of Statistics, Computer Science, and Mathematics, Public University of Navarre, Campus de Arrosadia, 31006 Pamplona, Spain.
E-mail: lola@unavarra.es

Abstract

Spatio-temporal disease mapping studies the distribution of mortality or incidence risks in space and its evolution in time, and it usually relies on fitting hierarchical Poisson mixed models. These models are complex for practitioners as they generally require adding constraints to correctly identify and interpret the different model terms. However, including constraints may not be straightforward in some recent software packages. This paper focuses on NIMBLE, a library of algorithms that contains among others a configurable system for Markov chain Monte Carlo (MCMC) algorithms. In particular, we show how to fit different spatio-temporal disease mapping models with NIMBLE making emphasis on how to include sum-to-zero constraints to solve identifiability issues when including spatio-temporal interactions. Breast cancer mortality data in Spain during the period 1990-2010 is used for illustration purposes. A simulation study is also conducted to compare NIMBLE with R-INLA in terms of parameter estimates and relative risk estimation. The results are very similar but differences are observed in terms of computing time.

Keywords: Disease mapping, identifiability, INLA, NIMBLE, spatio-temporal interactions, sum-to-zero constraints

1 Introduction

Spatio-temporal disease mapping models are essential in public health and epidemiology to describe the temporal evolution of geographical patterns of mortality or incidence risks. These models provide crucial information to make hypothesis about health inequalites and potential risk factors which in turn contribute to a better comprehension of the etiology of the disease under study.

Common risk measures, such as the standardized mortality ratio (SMR), may be highly variable in small areas or if the disease under study is rare. Consequently, models to smooth the risks and reduce variance become essential. A large extent of the research in disease mapping is based on Bayesian hierarchical spatio-temporal models that borrow strength from space and time to smooth the risks and reduce their variability. In particular, generalized linear mixed models (GLMMs) assuming a Poisson likelihood and a logarithmic link function have been widely adopted for smoothing purposes and inference. Research on spatio-temporal disease mapping models is now extensive, and alternative proposals for the different components (spatial, temporal and spatio-temporal interactions) of the

model are available. The spatial term has been generally approached using conditional autoregressive priors. In particular, the intrinsic conditional autoregressive (ICAR) prior (Besag, 1974) has been widely used, though other possibilities based on P-splines have been proposed (see for example Goicoa et al., 2012). For the temporal component, parametric (Bernardinelli et al., 1995; Assunção et al., 2001) as well as non parametric (Knorr-Held, 2000; Ugarte et al., 2014) or spline based (Ugarte et al., 2012) models have been proposed. Additionally, space-time interaction terms have been generally incorporated in the models to capture specificities of the small areas. Without being exhaustive, some examples include interactions terms based on Markov random fields (Knorr-Held, 2000) or on P-splines (Ugarte et al., 2010). In this paper we focus on spatio-temporal models including an intercept that in general can be interpreted as an overall risk level, and spatial, temporal, and space-time interaction random effects. More precisely we assume ICAR spatial random effects, temporal random effects modeled with a first order random walk (RW1), and space-time interaction terms defined as Markov random fields (Knorr-Held, 2000). It is known that these models present some identifiability issues that have been studied in the literature (see for example Gelfand and Sahu, 1999). These identifiability problems arise because the ICAR and the RW1 priors are improper and they have implicitly defined an intercept. Additionally, the interaction term overlaps with the main spatial and temporal random effects resulting in additional identifiability issues. Goicoa et al. (2018) provide a comprehensive study about this topic describing why the identifiability problems arise and how to deal with them through a reparameterization or the inclusion of sum-to-zero constraints in the estimation process.

Nowadays, there exist different free software packages to fit spatio-temporal models in disease mapping within a fully Bayesian approach. Here we focus on NIMBLE (de Valpine et al., 2017) and R-INLA (Lindgren and Rue, 2015) to compare both tools in terms of parameters and relative risk estimation. NIMBLE is a recent package that permits to fit hierarchical models using a configurable system of Markov chain Monte Carlo (MCMC) algorithms. The main characteristic of NIMBLE compared to other software packages that run MCMC algorithms is that it extends the BUGS language for writing new distributions and functions and therefore allowing for more flexible model specifications. NIMBLE is built in R but compiles the models and algorithms using C++ to speed up computations. Moreover, it provides its own language for writing new algorithms. R-INLA is an R-package for approximate Bayesian inference based on integrated nested Laplace approximations (INLA) (Rue et al., 2009) and numerical integration that has become popular during the last years due to its flexibility, computational time savings compare to MCMC algorithms, and because it avoids MCMC convergence problems.

The main goal of this work is to show how to fit well-known and easily interpretable spatio-temporal disease mapping models with NIMBLE. In particular, we pay special attention on how to include sum-to-zero constraints on the interaction term. Usually sum to zero constraints are introduced by centering the random effects in each iteration of the MCMC sampling scheme. However, this is generally very computationally intensive. Here we take advantage of one construction in NIMBLE to fit intrinsic conditional autoregressive models that constraints the random effects to sum to zero. This requires to express the interaction term as independent standard normal random variables and to include the spatial and temporal dependence through pre and post multiplication by appropriate matrices (see Martínez-Beneito, 2013). Additionally, we pursue to compare the performance of NIMBLE and R-INLA by means of an extensive simulation study. Both packages will be used to analyse breast cancer mortality in Spanish provinces during the period 1990-2010.

The rest of the paper is organized as follows. Section 2 introduces the class of spatio-temporal models that we will fit in this paper. Section 3 explains how to include sum-to-zero constraints in

NIMBLE appropriately. In Section 4, some characteristics of NIMBLE are briefly described. Section 5 is devoted to the analysis of the Spanish breast cancer data. In Section 6, a simulation study is conducted to compare R-INLA and NIMBLE. The paper closes with a discussion.

2 Spatio-temporal models in disease mapping

In this section we review the easily interpretable class of non-parametric spatio-temporal models including different types of space-time interactions proposed by Knorr-Held (2000).

Let us consider a large domain (e.g. a country) divided into S small areas (i.e. provinces) labelled as $i = 1, 2, ..., S$. For each small area i data are available at different time points t, $t = 1, 2, ..., T$. Let us also denote by Y_{it} the number of deaths (or incident cases) in the ith small area at time t. Then, conditional on the relative risk r_{it}, Y_{it} is assumed to be Poisson distributed with mean $\mu_{it} = e_{it} r_{it}$, where e_{it} represents the number of expected cases for area i at time t. That is

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

Here, the log risk, $\log r_{it}$, is modeled as

$$\log r_{it} = \alpha_0 + \xi_i + \gamma_t + \delta_{it}, \tag{1}$$

where α_0 can be interpreted as an overall risk, $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_S)'$ is the vector of spatial random effects and represents the underlying spatial pattern, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_T)'$ is the vector of temporal random effects describing the global temporal trend common to all small areas, and finally $\boldsymbol{\delta} = (\delta_{11}, ..., \delta_{S1}, ..., \delta_{1T}, ..., \delta_{ST})'$ is the vector of spatio-temporal random effects to capture the specificities of the small areas in each time point. In this paper, we consider two prior distributions for the spatial random effects. The first one is an ICAR prior (Besag, 1974), that is, the vector of spatial effects $\boldsymbol{\xi}$ follows the improper distribution with Gaussian kernel, $p(\boldsymbol{\xi}) \propto \exp\left(-\frac{1}{2\sigma_\xi^2} \boldsymbol{\xi}' \mathbf{Q}_\xi \boldsymbol{\xi}\right)$, where $\mathbf{Q}_\xi$ is the neighbourhood matrix defined as $\mathbf{Q}_{\xi(ij)} = -1$ if areas i and j are neighbours and 0 otherwise, and $\mathbf{Q}_{\xi(ii)}$ is equal to the number of neighbours of the ith region. Here, two regions are neighbours if they share a common border. The second prior distribution for the spatial random effects is the convolution prior proposed by Besag et al. (1991) (denoted by BYM hereafter in the paper). The BYM model includes two spatial random effects $\boldsymbol{\xi} = \mathbf{u} + \mathbf{v}$. The vector of random effect $\mathbf{u}$ captures spatially structured variability through an ICAR prior whereas the vector of random effects $\mathbf{v}$ deals with spatially unstructured heterogeneity using an exchangeable distribution. That is $p(\mathbf{u}) \propto \exp\left(-\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{Q}_\xi \mathbf{u}\right)$, and $p(\mathbf{v}) \propto \exp\left(-\frac{1}{2\sigma_v^2} \mathbf{v}' \mathbf{I}_\xi \mathbf{v}\right)$, where $\mathbf{I}_\xi$ is an $S \times S$ identity matrix. For the vector of temporal random effects $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_T)'$, a first order random walk (RW1) prior is assumed. Then, $p(\boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2\sigma_\gamma^2} \boldsymbol{\gamma}' \mathbf{Q}_\gamma \boldsymbol{\gamma}\right)$, where $\mathbf{Q}_\gamma$ is the RW1 structure matrix (see Rue and Held, 2005, page 95). Note that $\mathbf{Q}_\gamma$ is a "neighbourhood" matrix in time, where each time point (except the first and the last one) has two neighbours, the preceding and the subsequent one. Finally, for the vector of the interaction random effects $\boldsymbol{\delta} = (\delta_{11}, ..., \delta_{S1}, ..., \delta_{1T}, ..., \delta_{ST})'$, the following prior distribution with Gaussian kernel is assumed, $p(\boldsymbol{\delta}) \propto \exp\left(-\frac{1}{2\sigma_\delta^2} \boldsymbol{\delta}' \mathbf{R}_\delta \boldsymbol{\delta}\right)$. Here, four different types of interaction effects are considered (Knorr-Held, 2000) depending on the structure matrix $\mathbf{R}_\delta$. Namely, Type I interaction in which all the elements are independent ($\mathbf{R}_\delta = \mathbf{I}_\delta$, where $\mathbf{I}_\delta$ is an $ST \times ST$ identity matrix); in Type II interaction the terms are structured in time but not in space and $\mathbf{R}_\delta = \mathbf{Q}_\gamma \otimes \mathbf{I}_\xi$; Type III interaction deals with interaction terms structured in space but

not in time, then $\mathbf{R}_\delta = \mathbf{I}_\gamma \otimes \mathbf{Q}_\xi$, where $\mathbf{I}_\gamma$ is a $T \times T$ identity matrix. Lastly, in Type IV interaction the elements are structured in space and time and $\mathbf{R}_\delta = \mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi$.

The spatio-temporal model (1) presents some identifiability problems. First, the spatial ICAR and the temporal RW1 priors are improper and they have implicitly defined an intercept, consequently the model has three intercepts, the one explicitly included in the model, and those arising from the spatial and temporal priors. Second, the interaction term overlaps with the main spatial and temporal effects. This overlap depends on the type of interaction. Goicoa et al. (2018) provide a thorough insight on these identifiability issues depending on the interaction type and they indicate appropriate set of constraints on the random effects to identify the model. The readers are referred to this paper for full details. Here, Table 1 summarises the constraints needed to identify the different terms of the model for each type of interaction.

Table 1: Set of constraints on the random effects for the different type of interaction terms.

Interaction type	Spatial term	Temporal term	Interaction term
Type I	$\sum_{i=1}^{S} \xi_i = 0$	$\sum_{t=1}^{T} \gamma_t = 0$	$\sum_{t=1}^{T}\sum_{i=1}^{S} \delta_{it} = 0$
Type II	$\sum_{i=1}^{S} \xi_i = 0$	$\sum_{t=1}^{T} \gamma_t = 0$	$\sum_{t=1}^{T} \delta_{it} = 0,\ i = 1, \ldots, S$
Type III	$\sum_{i=1}^{S} \xi_i = 0$	$\sum_{t=1}^{T} \gamma_t = 0$	$\sum_{i=1}^{S} \delta_{it} = 0,\ t = 1, \ldots, T$
Type IV	$\sum_{i=1}^{S} \xi_i = 0$	$\sum_{t=1}^{T} \gamma_t = 0$	$\sum_{t=1}^{T} \delta_{it} = 0,\ i = 1, \ldots, S$ $\sum_{i=1}^{S} \delta_{it} = 0,\ t = 1, \ldots, T$

3 Fitting spatio-temporal models with NIMBLE

In this section we explain how to include sum-to-zero constraints to fit model (1) depending on the interaction type. More precisely, we comment on how to deal with the ICAR spatial prior and the RW1 temporal prior in NIMBLE. We pay special attention on the constraints needed for the interaction terms and provide an effective way of including the appropriate set of constraints taking advantage of existing functions in NIMBLE. Including constraints in R-INLA is relatively easy and the literature provides examples (see Goicoa et al., 2018, and the references therein). The full code and data to reproduce results will be avalaible at https://github.com/spatialstatisticsupna/Comparing-R-INLA-and-NIMBLE. R-code to fit spatio-temporal models with a Type IV interaction is provided in the Appendix.

3.1 Spatial ICAR and temporal RW1 priors

The spatial ICAR prior is implemented in NIMBLE through the function `dcar_normal`. This function allows the user to impose easily the sum-to-zero constraint $\sum_{i=1}^{S} \xi_i = 0$ specifying the argument `zero_mean = 1`, and providing a straightforward approach to identify the spatial term.

This function relies on the full conditionals $\xi_i|\boldsymbol{\xi}_{-i} \sim N\left(\frac{1}{n_i}\sum_{i\sim j}\xi_j, \frac{\sigma_\xi^2}{n_i}\right)$, where $\boldsymbol{\xi}_{-i}$ represents all elements of $\boldsymbol{\xi}$ excluding ξ_i, n_i is the number of neighbours of the ith region and $i \sim j$ indicates the set of neighbours of the ith region. This function is analogous to the `car.normal` function in GeoBUGS, an add-on to WinBUGS (Lunn et al., 2000). For more details about the `dcar_normal` function the reader is referred to the NIMBLE user manual (de Valpine et al., 2021).

The temporal RW1 prior can be included in NIMBLE in two alternative ways. The first one (denoted **Nimble 1** hereafter in the paper) consists of using the `dcar_normal` function given that the RW1 model is an ICAR model in time. Similar to the spatial case, the sum-to-zero constraint $\sum_{t=1}^{T}\gamma_t = 0$ is specified with the argument `zero_mean = 1`. The second possibility (denoted **Nimble 2**) implements the RW1 through the conditionals $\gamma_t|\gamma_{t-1} \sim N\left(\gamma_{t-1}, \sigma_\gamma^2\right)$, $t = 2, \ldots, T$ (Lawson, 2020). Note that this procedure does not need the sum-to-zero constraint $\sum_{t=1}^{T}\gamma_t = 0$ as we fix the first element of the temporal random effect equal to zero. This latter constraint identifies the temporal effect, however it cannot be interpreted as the global temporal trend. In fact, the temporal trend obtained with Nimble 2 is a shifted version of the one obtained with the Nimble 1 procedure and with R-INLA. To interpret the temporal component obtained with Nimble 2 as a global temporal trend, we should consider the sum $\alpha_0 + \gamma_t$ or calculate posterior patterns (Adin et al., 2017).

3.2 Spatio-temporal interactions in NIMBLE

Placing sum-to-zero constraints on the interaction term is not an easy task in NIMBLE. In general, algorithms may fail if they are not specially designed to work with constraints. One way to set the sum-to-zero constraints is to appropriately center (in space, in time or both) the random effects in each iteration of the MCMC scheme. However, this can be extremely inefficient. In this subsection we reorganize the vector of interaction random effects in matrix form so that we can use the `dcar_normal` function by rows or by columns and introduce the constraints through this function in type II and type III interactions. Moreover, we show the rationale to include constraints in the type IV interaction using the `dcar_normal` and multiplying by appropriate matrices.

Let us assume that the spatio-temporal interaction random effect $\boldsymbol{\delta} = (\delta_{11}, ..., \delta_{S1}, ..., \delta_{1T}, ..., \delta_{ST})'$ has the separable covariance structure (up to a scale parameter σ_δ^2) $\boldsymbol{\Sigma}_\gamma \otimes \boldsymbol{\Sigma}_\xi$, where $\boldsymbol{\Sigma}_\gamma$ and $\boldsymbol{\Sigma}_\xi$ are the temporal and spatial covariance matrices respectively. Let us rearrange the interaction vector into the $S \times T$ matrix

$$\boldsymbol{\delta}^* = \begin{pmatrix} \delta_{11} & \delta_{12} & \ldots & \delta_{1T} \\ \delta_{21} & \delta_{22} & \ldots & \delta_{2T} \\ \vdots & \vdots & & \vdots \\ \delta_{S1} & \delta_{S2} & \ldots & \delta_{ST} \end{pmatrix},$$

where each row represents a specific temporal trend for each area and each column represents a specific spatial pattern for each time point. Using ideas from the multivariate proposal of Martínez-Beneito (2013), the matrix of the interaction effects can be expressed as

$$\boldsymbol{\delta}^* = \widetilde{\boldsymbol{\Sigma}}_\xi \boldsymbol{\epsilon} \widetilde{\boldsymbol{\Sigma}}_\gamma', \tag{2}$$

where the tilde indicates the lower triangular matrix of the Cholesky decomposition of any symmetric and positive definite matrix, and $\boldsymbol{\epsilon}$ is a $S \times T$ matrix of independent standard normal random variables. Pre-multiplication by $\widetilde{\boldsymbol{\Sigma}}_\xi$ blends the different rows of $\boldsymbol{\epsilon}$ and induces spatial dependence.

Analogously, information from different columns of $\boldsymbol{\epsilon}$ is combined through post-multiplication by $\widetilde{\boldsymbol{\Sigma}}_\gamma$. Hence, the rows of $\boldsymbol{\delta}^*$ are structured in time and the columns are structured in space. Using the following result (see for example Harville, 2008, p. 345)

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B),$$

where the vec() operator puts the columns of a matrix one under the other, it follows that

$$\boldsymbol{\delta} = \text{vec}(\boldsymbol{\delta}^*) = (\widetilde{\boldsymbol{\Sigma}}_\gamma \otimes \widetilde{\boldsymbol{\Sigma}}_\xi)\text{vec}(\boldsymbol{\epsilon}),$$

and then

$$\text{cov}(\boldsymbol{\delta}) = \text{cov}(\text{vec}(\boldsymbol{\delta}^*)) = (\widetilde{\boldsymbol{\Sigma}}_\gamma \otimes \widetilde{\boldsymbol{\Sigma}}_\xi)\text{cov}(\text{vec}(\boldsymbol{\epsilon}))(\widetilde{\boldsymbol{\Sigma}}_\gamma \otimes \widetilde{\boldsymbol{\Sigma}}_\xi)' = \sigma_\delta^2(\boldsymbol{\Sigma}_\gamma \otimes \boldsymbol{\Sigma}_\xi),$$

taking into account that $\text{cov}(\text{vec}(\boldsymbol{\epsilon})) = \sigma_\delta^2\mathbf{I}_\delta$. It is worth noting that in our case $\boldsymbol{\Sigma}_\xi = \mathbf{Q}_\xi^-$ and $\boldsymbol{\Sigma}_\gamma = \mathbf{Q}_\gamma^-$, where the symbol $^-$ represents the Moore-Penrose generalized inverse of a matrix, are not positive definite matrices, but nonnegative definite. In fact, the ranks of these matrices are $r_\xi = S - 1$ (assuming a connected spatial graph) and $r_\gamma = T - 1$ respectively. However, there still exist unique lower triangular matrices $\widetilde{\boldsymbol{\Sigma}}_\xi$ and $\widetilde{\boldsymbol{\Sigma}}_\gamma$ with one null row such that $\boldsymbol{\Sigma}_\xi = \widetilde{\boldsymbol{\Sigma}}_\xi\widetilde{\boldsymbol{\Sigma}}_\xi'$ and $\boldsymbol{\Sigma}_\gamma = \widetilde{\boldsymbol{\Sigma}}_\gamma\widetilde{\boldsymbol{\Sigma}}_\gamma'$ (see Harville, 2008, Theorem 14.5.16, p. 234). Using this approach, we next explain how to include sum-to-zero constraints on the interaction terms depending on the type of interaction.

3.2.1 Type I interaction

In Type I interaction, Equation (2) reduces to $\boldsymbol{\delta}^* = \mathbf{I}_\xi\boldsymbol{\epsilon}\mathbf{I}_\gamma$, and there is neither spatial nor temporal structure. Here, the covariance matrix of the interaction term is $\sigma_\delta^2\mathbf{I}_\delta$, which is of full rank, and in this case there is a mild identifiability issue or a "confounding issue" with the intercept. That is the reason why one sum-to-zero constraint is needed. The way to include this constraint is to center the interaction random effect in each iteration of the MCMC scheme by subtracting the mean $\frac{1}{ST}\sum_{t=1}^{T}\sum_{i=1}^{S}\delta_{it}$ to each element δ_{it}. However, this form of including this constraint is fairly inefficient.

3.2.2 Type II interaction

In Type II interaction, Equation (2) takes the form $\boldsymbol{\delta}^* = \mathbf{I}\boldsymbol{\epsilon}\widetilde{\boldsymbol{\Sigma}}_\gamma' = \boldsymbol{\epsilon}\widetilde{\boldsymbol{\Sigma}}_\gamma'$. Consequently the terms are structured in time but not in space, that is $\text{cov}(\boldsymbol{\delta}) = \text{cov}(\text{vec}(\boldsymbol{\delta}^*)) = \sigma_\delta^2(\boldsymbol{\Sigma}_\gamma \otimes \mathbf{I}_\xi)$. In brief, the rows $\delta_{i\cdot}^* = (\delta_{i1}, \delta_{i2}, ..., \delta_{iT})'$, $i = 1, \ldots, S$ of the matrix $\boldsymbol{\delta}^*$ follow a RW1 distribution. An easy way to implement the Type II interaction in NIMBLE is assigning a RW1 distribution with the same precision parameter to each row of the matrix $\boldsymbol{\delta}^*$ using the `dcar_normal` function. Including the argument `zero_mean = 1`, the set of constraints $\sum_{t=1}^{T}\delta_{it} = 0$, $i = 1, 2, ..., S$ are automatically incorporated.

3.2.3 Type III interaction

Type III interaction is analogous to Type II, but now Equation (2) becomes $\boldsymbol{\delta}^* = \widetilde{\boldsymbol{\Sigma}}_\xi\boldsymbol{\epsilon}\mathbf{I}_\gamma' = \widetilde{\boldsymbol{\Sigma}}_\xi\boldsymbol{\epsilon}$. Hence the terms are structured in space but not in time, that is $\text{cov}(\boldsymbol{\delta}) = \text{cov}(\text{vec}(\boldsymbol{\delta}^*)) = \sigma_\delta^2(\mathbf{I}_\gamma \otimes \boldsymbol{\Sigma}_\xi)$. In this case each column $\delta_{\cdot t}^* = (\delta_{1t}, \delta_{2t}, ..., \delta_{St})'$, $t = 1, 2, ..., T$ of the matrix $\boldsymbol{\delta}^*$ follows an ICAR distribution. To implement this interaction type in NIMBLE we assign an ICAR prior to each

column of $\boldsymbol{\delta}^*$ using the `dcar_normal` function. Similarly to the Type II interaction, including the argument `zero_mean = 1` automatically incorporates the set of constraints $\sum_{i=1}^{S} \delta_{it} = 0$, $t = 1, \ldots, T$.

3.2.4 Type IV interaction

Type IV interaction is the most challenging interaction type to implement in NIMBLE. In Type IV, Equation (2) takes the form $\boldsymbol{\delta}^* = \widetilde{\boldsymbol{\Sigma}}_{\xi} \boldsymbol{\epsilon} \widetilde{\boldsymbol{\Sigma}}_{\gamma}'$ and the elements are structured in space and in time, that is $\text{cov}(\boldsymbol{\delta}) = \text{cov}(\text{vec}(\boldsymbol{\delta}^*)) = \sigma_{\delta}^2(\boldsymbol{\Sigma}_{\gamma} \otimes \boldsymbol{\Sigma}_{\xi})$. In this interaction type each row of $\boldsymbol{\delta}^*$ follows a RW1 distribution and each column has an ICAR distribution. To implement this interaction in NIMBLE one would be tempted to use the `dcar_normal` function by rows and columns, but this is not possible. The function can be used by rows or by columns, not both. Consequently, there are two possibilities to implement this interaction type in NIMBLE:

Option 1: The key point here is to express Equation (2) as $\boldsymbol{\delta}^* = \widetilde{\boldsymbol{\Sigma}}_{\xi} \boldsymbol{\epsilon} \widetilde{\boldsymbol{\Sigma}}_{\gamma}' = \boldsymbol{\phi} \widetilde{\boldsymbol{\Sigma}}_{\gamma}'$ where $\boldsymbol{\phi}$ is a $S \times T$ matrix whose columns are Gaussian Markov random fields in space with an ICAR distribution. Then, we assign a spatial ICAR prior to each column using the `dcar_normal` function. The argument `zero_mean = 1` automatically incorporates the set of constraints $\sum_{i=1}^{S} \delta_{it} = 0$, $t = 1, \ldots, T$. However, doing this the temporal dependence is not included in the interaction term and hence we have to postmultiply the set of ICAR columns by the matrix $\widetilde{\boldsymbol{\Sigma}}_{\gamma}'$. This matrix introduces the temporal dependence and the constraints $\sum_{t=1}^{T} \delta_{it} = 0$, $i = 1, 2, ..., S$ are automatically incorporated due to the rank deficiency of the matrix.

Option 2: Here, Equation (2) is expressed as $\boldsymbol{\delta}^* = \widetilde{\boldsymbol{\Sigma}}_{\xi} \boldsymbol{\epsilon} \widetilde{\boldsymbol{\Sigma}}_{\gamma}' = \widetilde{\boldsymbol{\Sigma}}_{\xi} \boldsymbol{\psi}$, where $\boldsymbol{\psi}$ is a $S \times T$ matrix whose rows are Gaussian Markov random fields in time with a RW1 distribution. We assign a RW1 prior to each row using the `dcar_normal` function with the argument `zero_mean = 1` to incorporate the constraints $\sum_{t=1}^{T} \delta_{it} = 0$, $i = 1, \ldots, S$. Similar to Option 1, the spatial dependence is not incorporated in the interaction term with this procedure, consequently we premultiply the set of RW1 rows by the matrix $\widetilde{\boldsymbol{\Sigma}}_{\xi}$ to incorporate the spatial correlation. As the spatial matrix is rank deficient, the constraints $\sum_{i=1}^{S} \delta_{it} = 0$, $t = 1, 2, ..., T$ are automatically incorporated.

Though both procedures are equivalent, Option 1 is preferred. The reason is that usually the number of areas S is larger than the number of time periods T. Then Option 2 is more time consuming because we assign a RW1 distribution to S rows and we premultiply by a $S \times S$ lower triangular matrix. Instead, in Option 1, we assign an ICAR distribution to T columns and we postmultiply by a $T \times T$ upper triangular matrix. In our particular example with $S = 50$ and $T = 21$, Option 2 is about 25 times slower than Option 1.

4 MCMC methods in NIMBLE

NIMBLE is a library of algorithms that contains MCMC, Monte Carlo expectation maximization (MCEM), and Sequential Monte Carlo algorithms among others (the last one is implemented in another package called nimbleSMC (de Valpine et al., 2021)). For MCMC algorithms, NIMBLE

assigns a default sampler to each parameter (or block of parameters) in the model and allows to replace some or all of them by other samplers already implemented or by a specific sampler written by the user.

To fit the spatio-temporal models introduced in this paper with NIMBLE, MCMC algorithms with default samplers are used because the posterior samples of the parameters mix well and do not show convergence problems. In particular, a Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; Casella and George, 1992) is assigned to the random effects with an ICAR prior distribution. The main advantage of Gibbs sampling is that instead of sampling from a complex joint posterior distribution it reduces the simulation process to a sequence of algorithms for sampling from a one or low dimensional distribution. Regarding the hyperparameters, NIMBLE assigns by default Metropolis-Hastings adaptive random-walk sampler (Roberts and Sahu, 1997). This sampler is also used for other random effects with a prior other than the ICAR, such as the temporal random effects defined as in Nimble 2 and type I space-time interaction. If the Metropolis-Hastings adaptive random-walk sampler shows convergence problems, an alternative might be to use a slice sampler (Neal, 2003). In general, mixing with slice sampling might be better, but the computational cost is high compared to Metropolis-Hastings adaptive random-walk samplers.

To go deeper with the samplers implemented in NIMBLE and to write customized samplers, the reader is referred to the user manual (de Valpine et al., 2021).

5 Spanish breast cancer mortality data

In this section we illustrate the fit of the spatio-temporal model (1) with the four interaction types with NIMBLE and compare the results with the R-INLA fit. We analyse female breast cancer mortality data (ICD-10 code 50) in Spain during the period 1990-2010.

The procedures in Section 3 are used to consider the ICAR and BYM priors for the spatial component, the RW1 prior for the temporal random effects, and the different type of interactions. Three MCMC chains have been run for each model with 30000 iterations each discarding the first 5000 as a burn-in period. One out of every 75 iterations is saved leading to a total of 999 iterations. Convergence of MCMC samples is assessed through the generalized method of Gelman-Rubin (Brooks and Gelman, 1998) and the effective sample size (see for example Gelman et al., 2013, Chapter 17) which are implemented in R package `coda` (Plummer et al., 2006). This package contains a wide range of methods to assess convergence of MCMC samples. In addition, graphical checks of chains and their autocorrelations were performed to assess convergence. To check that the simulated sample size is sufficient to achieve a specific level of accuracy we have compared the Monte Carlo standard error (MCSE) with the estimated posterior standard deviation of the parameters. Usually chains are run until MCSE is less than 5% of the posterior standard deviation, see for example the WinBUGS manual (Lunn et al., 2000; Spiegelhalter et al., 2003). In R-INLA all models have been fitted using the full Laplace approximation.

Prior distributions of the hyperparameters may have an impact on the final results. Here, as recommended in Gelman (2006), vague uniform priors on the standard deviations, $\sigma_\xi, \sigma_\gamma, \sigma_\delta$ are considered in NIMBLE and R-INLA. This set of hyperprior distributions is denoted as **H1**. In order to detect some sensitivity to the hyperpriors, the models have been fitted using two additional sets of hyperpriors. More precisely, a Gamma(1,0.00005) on all the precision parameters $\tau_\xi = 1/\sigma_\xi^2, \tau_\gamma = 1/\sigma_\gamma^2, \tau_\delta = 1/\sigma_\delta^2$ (the default prior in R-INLA for the ICAR, BYM and RW1) denoted by **H2**. Finally, we also consider a Gamma(1, 0.01) prior on the spatial precision parameter τ_ξ

and a Gamma(1,0.00005) prior on the temporal and interaction precision parameters τ_γ and τ_δ, denoted as **H3**. A total of 48 models have been fitted in NIMBLE depending on the spatial random effect (ICAR or BYM), the way of implementing the RW1 (Nimble 1 or Nimble 2), the four types of interaction, and the three set of hyperpriors. In R-INLA, 24 models are fitted as we only consider one way to implement the RW1 through the "rw1" model available in the package. To select the best candidate, we look at two model selection criteria, the Deviance Information Criterion, DIC (Spiegelhalter et al., 2002), and the Watanabe-Akaike Information Criterion, WAIC (Watanabe, 2010). The DIC is not directly available in NIMBLE, so it has been manually implemented (see the code available in the Appendix).

Table 2 displays the values of the DIC and the WAIC together with the mean deviance ($\overline{D(\theta)}$) and the effective number of parameters (p_D) for the spatio-temporal models with the ICAR and BYM spatial priors, the two forms of including the RW1 in NIMBLE, Nimble 1 and Nimble 2, the four types of interaction, and the set of hyperpriors **H1** (results for hyperpriors **H2** and **H3** are shown in Tables A.1 and A.2 respectively in the Supplementary Materials). In general, DIC and WAIC values that differ by less than five point towards the same model (see for example Vranks et al., 2021). According to these criteria, the ICAR and BYM spatial priors lead to very similar results, and the Type IV interaction is preferred over the others. Therefore, in what follows, we will display results for the Type IV interaction. Finally, according to these criteria, there are very minor differences between the alternative ways to include the RW1 in NIMBLE.

Although here DIC and WAIC calculated with R-INLA and NIMBLE show very small differences and clearly point towards the model with a Type IV interaction, Vranckx et al. (2021) warn about disparities among model selection criteria values provided by different software packages. Differences may be due to software-specific posterior samples or software-specific calculations of the model selection statistics.

Table 2: Mean deviance ($\overline{D(\theta)}$), effective number of parameters (p_D), DIC and WAIC for spatio-temporal models with an ICAR and a BYM spatial prior.

		ICAR spatial prior				**BYM spatial prior**			
		$\overline{D(\theta)}$	p_D	DIC	$WAIC$	$\overline{D(\theta)}$	p_D	DIC	$WAIC$
Type I	R-INLA	7593.1060	224.5179	7817.6239	7822.0751	7592.6220	225.4980	7818.1199	7821.6368
	Nimble 1	7598.0052	221.5838	7819.5891	7826.3674	7597.4211	221.8959	7819.3170	7826.4275
	Nimble 2	7598.5300	221.4272	7819.9570	7825.3750	7594.9590	224.4277	7819.3870	7825.2970
Type II	R-INLA	7546.7413	165.8487	7712.5900	7714.2766	7545.5545	166.8645	7712.4190	7713.6589
	Nimble 1	7548.9699	166.0982	7715.0681	7717.5550	7547.2755	166.1233	7713.3987	7715.9592
	Nimble 2	7550.8400	164.6176	7715.4580	7718.4830	7546.9790	165.9412	7712.9200	7715.0020
Type III	R-INLA	7600.7795	190.0447	7790.8242	7799.4310	7599.6856	189.6805	7789.3661	7798.8774
	Nimble 1	7603.1994	189.3856	7792.5850	7802.3154	7603.8188	189.4632	7793.2820	7803.9177
	Nimble 2	7603.8980	189.1573	7793.0550	7803.9930	7600.5790	189.8335	7790.4130	7799.8380
Type IV	R-INLA	7555.5723	148.8464	**7704.4186**	**7706.3231**	7554.4509	149.8304	**7704.2813**	**7705.7522**
	Nimble 1	7558.2180	148.0318	**7706.2498**	**7709.1370**	7556.3648	148.5247	**7704.8895**	**7708.0454**
	Nimble 2	7557.6240	149.8634	**7707.4880**	**7711.3580**	7557.2950	149.1446	**7706.4400**	**7709.4450**

To inspect the effect of the hyperprior distributions on the parameter estimates, Table 3 displays the posterior means and standard deviations of the intercept and the hyperparameters for a model with the ICAR spatial prior, the Type IV interaction and the three sets of hyperprior distributions. Little sensitivity in the parameter estimates is observed with regard to the prior distributions of the hyperparameters. The most remarkable element is the different posterior mean and standard deviation of the intercept when Nimble 2 is considered to implement the RW1 prior for the temporal effect. This is expected as the constraints used to identify the temporal effects are different to those used in Nimble 1 and R-INLA. In particular, the standard deviation of the Nimble 2 intercept is about three times higher than those obtained with R-INLA and Nimble 1. The results with the

BYM spatial prior are in line with those obtain with the ICAR spatial prior (see Table A.3 in the Supplementary Materials). Again, the most remarkable difference is found in the posterior mean of the intercept obtained with Nimble 2, though the posterior standard deviations of the intercept have been increased with the three methods. Some differences between Nimble 1 and Nimble 2 and R-INLA are observed in the estimation of the standard deviation of the spatially structured and unstructured spatial random effects. This is expected as in the BYM model only the sum $\sigma_u^2 + \sigma_v^2$ is identifiable. According to MacNab (2014), "no amount of data enables identification of σ_u^2 and σ_v^2 simultaneously".

Table 3: Posterior means and standard deviations of the intercept and the hyperparameters of the models with the ICAR spatial prior, Type IV interaction and the three sets of hyperprior distributions.

		H1		**H2**		**H3**	
		Mean	SD	Mean	SD	Mean	SD
α_0	R-INLA	-0.0349	0.0042	-0.0345	0.0040	-0.0346	0.0042
	Nimble 1	-0.0351	0.0041	-0.0346	0.0039	-0.0345	0.0039
	Nimble 2	0.0586	0.0137	0.0565	0.0129	0.0573	0.0125
σ_ξ	R-INLA	0.2011	0.0233	0.1933	0.0219	0.1946	0.0220
	Nimble 1	0.2012	0.0263	0.1931	0.0221	0.1949	0.0225
	Nimble 2	0.2001	0.0247	0.1927	0.0218	0.1950	0.0223
σ_γ	R-INLA	0.0246	0.0052	0.0219	0.0043	0.0219	0.0043
	Nimble 1	0.0246	0.0056	0.0221	0.0047	0.0222	0.0047
	Nimble 2	0.0233	0.0051	0.0223	0.0044	0.0219	0.0044
σ_δ	R-INLA	0.0389	0.0046	0.0374	0.0045	0.0373	0.0045
	Nimble 1	0.0387	0.0045	0.0374	0.0045	0.0374	0.0044
	Nimble 2	0.0392	0.0047	0.0376	0.0047	0.0372	0.0044

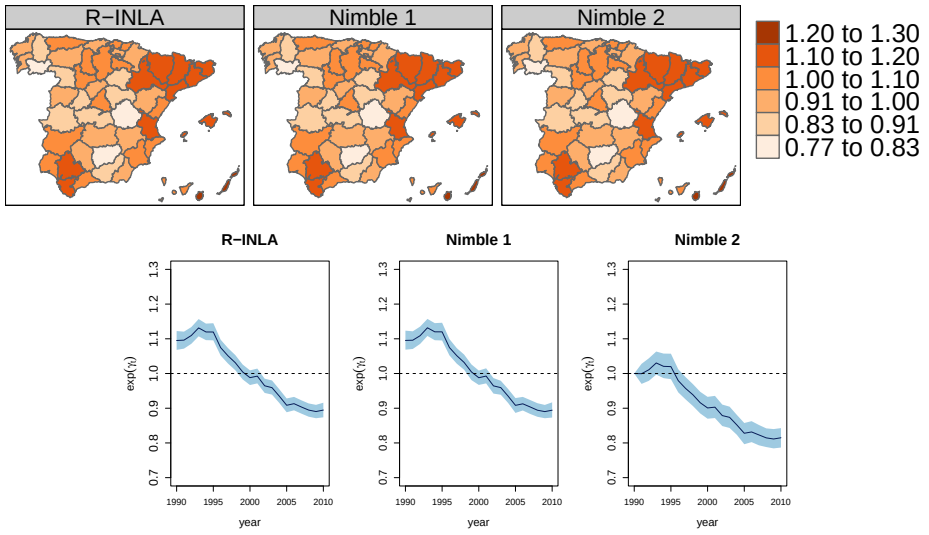

Figure 1: Spatial patterns and temporal trends (posterior means of $\exp(\xi_i)$ and $\exp(\gamma_t)$) estimated with R-INLA, Nimble 1 and Nimble 2 for ICAR model with H1 hyperpriors.

Given that there are no important differences between the models with the ICAR and BYM spatial priors, we will focus on the former. Figure 1 displays the spatial (top) and temporal (bottom) patterns obtained with R-INLA, Nimble 1, and Nimble 2. The spatial patterns in the three cases

are identical. Regarding the global temporal trend, R-INLA and Nimble 1 provide identical results, unlike Nimble 2 that provides a shifted temporal trend. This is expected as different constraints are used. While in R-INLA and Nimble 1 the temporal effects are centered, in Nimble 2 they are not. Consequently, the Nimble 2 temporal trend cannot be interpreted as the global temporal trend of the whole region. The different constraints have also an effect on the posterior standard deviation, and the credible intervals for the temporal effects are much wider with Nimble 2. In general, linear dependence of the random effects with the intercept leads to wider credible intervals, hence constraints centering the effects are preferred (Wood et al., 2013). To obtain the global temporal trend with Nimble 2, posterior patterns should be computed (Adin et al., 2017). Additionally, we could directly compare the temporal trends obtained with the three methods representing $\exp(\alpha_0 + \gamma_t)$ (see Figure A.1 in the Supplementary Materials). In spite of these differences in the temporal trend, the estimated relative risks are equal. Figure 2 displays dispersion plots of the relative risk estimates (posterior means) obtained with R-INLA vs. Nimble 1 (left) and Nimble 2 vs Nimble 1 (right) revealing identical estimates. Finally, Figure A.2 in the Supplementary Materials shows maps of the Type IV space-time interaction random effects (posterior mean of $\exp(\delta_{it})$) for years 1990, 1994, 1998, 2002, 2006 and 2010 obtained with R-INLA, Nimble 1 and Nimble 2. Temporal trends of the Type IV space-time interactions estimated with Nimble 1 and their corresponding credibility bands for some neighbouring provinces in the north and some other neighbouring provinces in the south of Spain are displayed in Figure A.3 in the Supplementary Materials. This figure reveals that neighbouring regions have similar trends (that is what a Type IV interaction means).

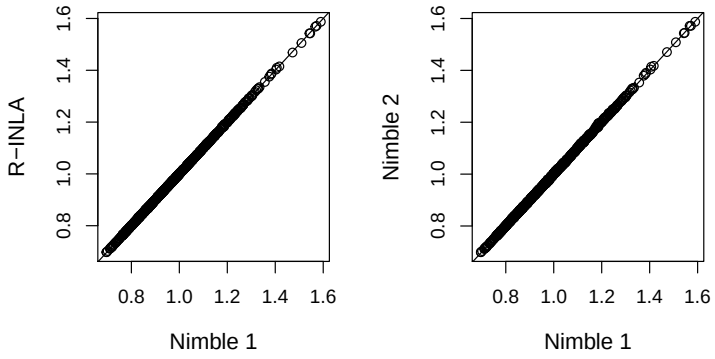

Figure 2: Dispersion plots of relative risk estimates (posterior means) obtained with R-INLA vs Nimble 1 and Nimble 2 vs Nimble 1 in the spatio-temporal model with the ICAR spatial prior, the Type IV interaction and the set of hyperpriors H1.

To close this section, it is worth remarking that R-INLA is faster than NIMBLE. More precisely, the computing time to fit the model with the ICAR spatial prior with INLA is 90.39 seconds (full Laplace) whereas in NIMBLE the computing time is 722.34 seconds with Nimble 1 and Nimble 2. The R-INLA and NIMBLE fits were run on a TTL personal computer with a 3.00 GHz Intel(R) Core(TM) i5-9500 CPU processor and 20GB RAM using R (version 4.0.3). INLA (version 21.02.23, dated 2021-02-22) and NIMBLE (version 0.11.1, dated 2021-05-23) were used.

6 Simulation study

In this section a simulation study is conducted for a thorough comparison between NIMBLE an R-INLA using the geographical set up of the Spanish breast cancer data. The simulation study is

focused on the spatio-temporal model with the ICAR spatial prior and the Type IV interaction.

A total of 500 data sets have been simulated according to the following scheme. For each data set $l = 1, 2, ..., 500$ the logarithm of relative risks is simulated as

$$\log r_{it}^l = \hat{\alpha_0} + \xi_i^l + \gamma_t^l + \delta_{it}^l$$

where $\hat{\alpha_0} = -0.0350$.

The random effects are generated from the following distributions with Gaussian kernels

$$p(\boldsymbol{\xi}) \propto \exp\left(-\frac{1}{2\hat{\sigma}_\xi^2}\boldsymbol{\xi}'\mathbf{Q}_\xi\boldsymbol{\xi}\right)$$
$$p(\boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2\hat{\sigma}_\gamma^2}\boldsymbol{\gamma}'\mathbf{Q}_\gamma\boldsymbol{\gamma}\right)$$
$$p(\boldsymbol{\delta}) \propto \exp\left(-\frac{1}{2\hat{\sigma}_\delta^2}\boldsymbol{\delta}'(\mathbf{Q}_\gamma \otimes \mathbf{Q}_\xi)\boldsymbol{\delta}\right)$$

where $\hat{\sigma}_\xi^2 = 0.2011^2$, $\hat{\sigma}_\gamma^2 = 0.0246^2$, and $\hat{\sigma}_\delta^2 = 0.0389^2$ are the square of the posterior means of $\sigma_\xi, \sigma_\gamma$ and σ_δ. Finally, the counts are simulated using a Poisson distribution $Y_{it}^l \sim Poisson(e_{it}r_{it}^l)$, where e_{it} are the number of expected cases in the Spanish breast cancer data. To examine the effect of the population size, the simulation study has been repeated multiplying the expected cases by the scale factors (SF) 0.5 and 2. For each simulated data set, the spatio-temporal model with the ICAR spatial prior and a Type IV interaction has been fitted with R-INLA, Nimble 1 and Nimble 2, and with the set of hyperpriors H1 and H2 defined in Section 5.

To assess how R-INLA, Nimble 1 and Nimble 2 recover the true value of the parameters used in the simulation, Table 4 shows the average over the 500 simulated data sets of the posterior means of the parameters for the three scenarios (SF=0.5, SF=1, and SF=2). In general, the three procedures recover the true value of the parameters regardless the scale factor on the expected cases. Interestingly, Nimble 2 also recovers the true value of the intercept on average (though the relative bias is about 7 times higher than with R-INLA and Nimble 1). This seems contradictory with the estimated value in the real analysis of the Spanish breast cancer data, where the posterior mean of the intercept is very different from those obtained with R-INLA and Nimble 1. The reason is that the variability in the Nimble 2 estimates is very large. Figure A.4 in the Supplementary Materials provides boxplots of the posterior means of the intercept obtained with the different methods in the 500 simulated data sets. Clearly, the variability in the Nimble 2 estimates is much larger.

Table 5 displays both simulated standard errors (sim) and estimated standard errors (est) multiplied by 100 to better interpret the results. The simulated standard error is defined as

$$\sqrt{\frac{1}{500}\sum_{l=1}^{500}(\hat{\theta}_l - \bar{\hat{\theta}})^2}$$

where $\hat{\theta}_l$ is the posterior mean of the parameter of interest in the l-th simulated dataset and $\bar{\hat{\theta}}$ is the average of the posterior means of the parameter computed over the 500 datasets. The simulated

Table 4: Mean values of estimated parameters with R-INLA, Nimble 1 and Nimble 2 for 500 simulated datasets. The number of expected cases are multiplied by scale factors (SF) 0.5, 1, and 2.

			R-INLA			**Nimble 1**			**Nimble 2**		
		True value	SF=0.5	SF=1	SF=2	SF=0.5	SF=1	SF=2	SF=0.5	SF=1	SF=2
α_0	H1	-0.0350	-0.0351	-0.0348	-0.0352	-0.0351	-0.0348	-0.0352	-0.0309	-0.0337	-0.0320
	H2		-0.0348	-0.0347	-0.0352	-0.0348	-0.0347	-0.0351	-0.0312	-0.0335	-0.0317
σ_ξ	H1	0.2011	0.2035	0.2049	0.2071	0.2022	0.2036	0.2057	0.2028	0.2044	0.2064
	H2		0.1953	0.1974	0.2001	0.1949	0.1975	0.2002	0.1950	0.1975	0.2003
σ_γ	H1	0.0250	0.0274	0.0264	0.0263	0.0270	0.0260	0.0258	0.0272	0.0261	0.0259
	H2		0.0222	0.0226	0.0233	0.0224	0.0228	0.0235	0.0224	0.0228	0.0235
σ_δ	H1	0.0390	0.0392	0.0395	0.0391	0.0393	0.0395	0.0390	0.0393	0.0395	0.0390
	H2		0.0362	0.0379	0.0381	0.0364	0.0381	0.0383	0.0365	0.0381	0.0383

standard error is the sample standard deviation of the estimates and can be interpreted as the true variability of the estimators. The estimated standard error is defined as

$$\frac{1}{500}\sum_{l=1}^{500} sd(\theta_l).$$

where $sd(\theta_l)$ is the posterior standard deviation of the parameter of interest. A large value of the simulated standard error indicates a great variability within the estimates of the parameter of interest. Estimated standard errors smaller (larger) than the simulated standard errors indicate underestimation (overestimation) of the standard errors. The results in Table 5 reveals a large simulated standard error for the intercept with Nimble 2 (as shown in Figure A.4 in the Supplementary Materials). This is probably the reason that the average of the posterior means over the 500 data sets is close to the true value but in the real analysis we obtain a different estimate in comparison with R-INLA and Nimble 1. As expected, the standard errors both simulated and estimated, decrease when the number of expected cases increases (see SF=0.5, SF=1, and SF=2).

Table 6 displays the empirical coverages of the credible intervals for the different parameters at nominal values 90%, 95% and 99%. In general, coverage rates are close to the nominal values with the three methods and regardless the scale factor when using hyperpriors H1 (which seems to be a better choice than H2). The exception is the coverage rates of the intercept obtained with Nimble 2, which is far from the nominal value. This is probably due to the high variability of the posterior means and the underestimation of the standard error (true simulated error and estimated error in Table 5). Additionally, with Nimble 2 the coverage rate for the intercept decreases when the SF increases. Though it seems contradictory, the reason for this is that the true simulated standard error decrease about a 3% from SF=0.5 to SF=1, and from SF=1 to SF=2, but the estimated standard error decreases about a 25% from SF=0.5 to SF=1, and from SF=1 to SF=2. Note also that in Nimble 2, the intercept cannot longer be interpreted as the overall level of risk, which is the meaning in the simulated values.

Mean values of the length of credibility intervals of relative risks and temporal effects ($\exp(\gamma_t)$) are computed based on 500 simulated data sets (not shown here to save space). In general, credibility intervals of $\exp(\gamma_t)$ are much wider for Nimble 2 than for R-INLA or Nimble 1. This agrees with the results in the Spanish breast cancer data (see Figure 1). However, this does not affect the estimates and width of the credibility intervals for the relative risks.

To measure the accuracy and precision of the relative risk estimates obtained with R-INLA, Nimble

Table 5: Simulated standard errors (sim) and estimated standard errors (est) for the different parameters obtained with R-INLA, Nimble 1 and Nimble 2. The number of expected cases are multiplied by scale factor (SF) 0.5, 1, and 2. Values are multiplied by 100.

R-INLA

		SF=0.5		SF=1		SF=2	
		sim	est	sim	est	sim	est
α_0	H1	0.5344	0.5459	0.3966	0.3945	0.2871	0.2990
	H2	0.5393	0.5451	0.3975	0.3950	0.2894	0.3124
σ_ξ	H1	2.3892	2.4313	2.2654	2.2901	2.1394	2.2189
	H2	2.3213	2.2843	2.2065	2.1511	2.0782	2.0840
σ_γ	H1	0.6787	0.7537	0.5810	0.6246	0.5064	0.5461
	H2	0.6520	0.6097	0.5579	0.5237	0.4782	0.4660
σ_δ	H1	0.6254	0.6156	0.4566	0.4644	0.3495	0.3594
	H2	0.6811	0.6042	0.4686	0.4559	0.3538	0.3541

Nimble 1

α_0	H1	0.5352	0.5464	0.3979	0.3910	0.2852	0.2797
	H2	0.5354	0.5445	0.3977	0.3926	0.2847	0.2798
σ_ξ	H1	2.3763	2.4792	2.2533	2.3437	2.1309	2.2763
	H2	2.3438	2.3121	2.2054	2.1806	2.0842	2.1218
σ_γ	H1	0.6652	0.7554	0.5742	0.6266	0.4976	0.5488
	H2	0.6523	0.6229	0.5558	0.5371	0.4748	0.4801
σ_δ	H1	0.6262	0.6249	0.4541	0.4677	0.3477	0.3610
	H2	0.6827	0.6123	0.4683	0.4614	0.3557	0.3587

Nimble 2

α_0	H1	6.4257	1.8595	6.2946	1.4019	6.1344	1.0361
	H2	6.4199	1.9165	6.2212	1.3653	6.0976	1.0198
σ_ξ	H1	2.3904	2.4760	2.2654	2.3444	2.1238	2.2656
	H2	2.3404	2.3098	2.2033	2.1833	2.0718	2.1208
σ_γ	H1	0.6676	0.7690	0.5736	0.6375	0.5019	0.5537
	H2	0.6964	0.6428	0.5551	0.5358	0.4819	0.4793
σ_δ	H1	0.6257	0.6258	0.4540	0.4682	0.3486	0.3610
	H2	0.6875	0.6140	0.4683	0.4601	0.3561	0.3571

Table 6: Empirical coverage probabilities of the parameters obtained with R-INLA, Nimble 1 and Nimble 2. The number of expected cases are multiplied by scale factors (SF) 0.5, 1, and 2.

						SF=0.5				
		R-INLA			**Nimble 1**			**Nimble 2**		
		90%	95%	99%	90%	95%	99%	90%	95%	99%
α_0	H1	91.2	95.8	99.2	91.0	95.4	98.4	35.6	41.6	53.6
	H2	89.4	95.0	99.2	90.6	95.0	98.8	35.4	41.2	52.4
σ_ξ	H1	90.6	95.8	99.2	90.2	96.0	99.2	90.4	95.6	99.0
	H2	88.4	93.2	98.2	87.6	92.6	98.6	87.6	93.0	98.6
σ_γ	H1	90.0	95.4	98.8	90.6	95.2	99.2	91.0	95.4	98.4
	H2	81.6	90.0	97.8	83.2	90.6	98.0	82.0	90.0	97.6
σ_δ	H1	91.0	95.8	98.2	90.8	95.6	98.8	91.2	95.2	98.6
	H2	82.4	91.4	98.8	82.4	90.2	98.2	84.0	91.6	96.8
						SF=1				
α_0	H1	88.2	93.8	99.6	88.2	93.0	98.6	27.0	31.2	40.2
	H2	88.6	94.2	99.6	88.6	93.6	99.2	27.2	31.2	41.4
σ_ξ	H1	88.6	93.6	98.6	89.4	94.8	98.8	88.4	93.2	99.2
	H2	87.0	93.2	98.0	87.2	93.2	98.4	88.2	92.6	98.8
σ_γ	H1	91.8	95.6	99.4	91.6	95.6	99.0	92.6	96.0	98.8
	H2	83.2	91.2	97.8	84.6	91.6	96.8	83.8	91.2	97.0
σ_δ	H1	89.2	93.6	98.4	89.2	94.4	98.4	90.2	94.2	98.4
	H2	87.0	92.6	98.0	87.8	93.0	97.6	87.4	92.0	97.4
						SF=2				
α_0	H1	89.6	95.6	99.8	87.8	94.2	98.6	20.6	24.6	33.2
	H2	91.2	97.0	99.4	88.0	94.6	98.8	20.2	25.0	32.8
σ_ξ	H1	90.2	95.0	98.8	90.0	96.2	99.4	91.8	95.2	99.0
	H2	89.8	95.2	98.6	89.6	94.6	99.0	89.8	95.0	98.6
σ_γ	H1	90.2	94.4	99.4	91.6	94.6	99.2	90.6	94.6	98.8
	H2	86.2	92.0	97.2	87.6	92.8	97.2	87.6	93.2	97.0
σ_δ	H1	91.4	95.8	99.2	91.0	96.2	99.2	91.2	96.2	99.2
	H2	90.0	95.2	98.4	90.6	95.2	98.4	90.2	95.0	98.6

1, and Nimble 2, the mean absolute relative bias (MARB) and the mean relative root mean squared error (MRRMSE) are calculated using the following expressions

$$MARB = \frac{1}{ST}\sum_{i=1}^{S}\sum_{t=1}^{T}\frac{1}{500}\left|\sum_{l=1}^{500}\frac{\hat{r}_{it}^{l}-r_{it}^{l}}{r_{it}^{l}}\right|,$$

$$MRRMSE = \frac{1}{ST}\sum_{i=1}^{S}\sum_{t=1}^{T}\sqrt{\frac{1}{500}\sum_{l=1}^{500}\left(\frac{\hat{r}_{it}^{l}-r_{it}^{l}}{r_{it}^{l}}\right)^{2}}.$$

Results of MARB and MRRMSE are shown in Table 7. Overall, all methods provide unbiased relative risk estimates, being the MARB less than 1% in all cases. The MRRMSE is pretty similar for the three methods and it is in general bellow 6%. As expected, both the MARB and the MRRMSE decrease when the number of expected cases increases. When the number of expected cases decrease (SF=0.5), the set of hyperprior distributions H1 provides slightly lower MARB and MRRMSE.

Table 7: Average value of mean absolute relative bias (MARB) and mean relative root mean prediction error (MRRMSE) of the relative risks estimated by R-INLA, Nimble 1 and Nimble 2 based on 500 simulated data sets.

		R-INLA			**Nimble 1**			**Nimble 2**		
		SF=0.5	SF=1	SF=2	SF=0.5	SF=1	SF=2	SF=0.5	SF=1	SF=2
MARB	H1	0.0021	0.0018	0.0013	0.0028	0.0022	0.0015	0.0028	0.0022	0.0015
	H2	0.0035	0.0018	0.0013	0.0030	0.0023	0.0015	0.0030	0.0023	0.0015
MRRMSE	H1	0.0502	0.0408	0.0342	0.0503	0.0409	0.0342	0.0503	0.0409	0.0342
	H2	0.0650	0.0409	0.0342	0.0504	0.0410	0.0342	0.0505	0.0410	0.0352

7 Discussion

The availability of high quality registers recording areal count data in time has promoted both the research on methodology and the development of new software to implement the theoretical procedures. In this work we focus on one of the more recently available software packages, NIMBLE, and we compare it with the well tested and widely used R-INLA. The latter provides approximate Bayesian inference using integrated nested Laplace approximation and numerical integration speeding up computations with regard to MCMC algorithms. Additionally, R-INLA handles constraints relatively easy and intuitively. A potential limitation is that it is designed for latent Gaussian Markov random fields, though many applications in practice can be described with this class of models. NIMBLE relies on MCMC algorithms to fit the models. Its strengths are that it extends the BUGS language and allows users to write their own models and customize the MCMC algorithms. However, regarding the spatio-temporal models considered in this paper, it is not straightforward to fit the different types of spatio-temporal interactions, with the exception of Type I (Lawson, 2020), as including the right constraints to overcome identifiability issues is not easy.

In this paper, we propose a simple way to impose sum-to-zero constraints in NIMBLE to identify spatio-temporal models including the ICAR and BYM spatial random effects, temporal random effects with a RW1 prior distribution, and the four types of interaction defined by Knorr-Held (2000). We take advantage of the functions already implemented in NIMBLE without modifying any existing algorithm. In particular, we have studied two alternative procedures to implement the

RW1 prior: one based on the full conditionals (Nimble 1) and the other one based on conditioning on the preceding time point (Nimble 2). Regarding the interaction, our approach rearranges the interaction terms in matrix form and introduces the spatial and temporal dependence by pre and post multiplying by appropriate matrices. Using this procedure, sum-to-zero constraints in models incorporating Type II, III, and IV interactions can be easily included. In the Type I interaction, the interaction term does not present an identifiability issue (as the covariance matrix is of full rank) but a confounding issue with the intercept. In this case, the sum-to-zero constraint is achieved by centering the interaction random effects in each iteration of the MCMC scheme. However this is computationally very demanding and the user may ignore this constraint. We have run the Type I interaction without constraints and the results are practically identical.

The procedures have been used to analyse female breast cancer data in Spanish provinces during the period 1990-2010 and we would like to remark interesting findings. First, the Nimble 2 option to implement the RW1 prior does not center the temporal trend, hence it cannot be interpreted as the global trend of the whole region, and the posterior mean of the intercept is also different to the estimates provided by Nimble 1 and R-INLA. Additionally, the credible bands are wider than those obtained with Nimble 1 and R-INLA. In general, linear dependence of the random effects with the intercept leads to wider credible intervals, hence constraints centering the effects are preferred (Wood et al., 2013). That is the reason why Nimble 2 leads to a poorer inference on the temporal trend. Second, the standard deviation parameter of the unstructured spatial term in the BYM spatial prior shows some sensitivity to the hyperprior distribution. This is expected as the data only identifies the sum of the variance parameters (structured+unstructured).

We have run a simulation study to compare the performance of NIMBLE and R-INLA. In general, Nimble 1, Nimble 2, and R-INLA recover the true values of the parameters used to generate the data. However, Nimble 2 option shows a large variability within the estimates (posterior means) of the intercept in each simulation run. This is also translated into a low coverage rate of the credible interval for this parameter. Overall, we have observed that Nimble 1 and R-INLA attain coverage rates for the parameters close to the nominal values regardless the scale factor multiplying the number of expected cases used in the simulation. Some sensitivity was observed with respect to the hyperprior distributions. Coverage rates improved when using vague uniform priors on the standard deviations as it was recommended by Gelman (2006). Finally, the procedures estimate the relative risk pretty well as the mean absolute relative bias is very low (less than 1%) in all cases and the mean relative mean squared error is below 6%. Both the bias and the error reduce when the number of expected cases increases.

To sum up, NIMBLE is a package that provides flexibility to include new models and customize algorithms, but these tasks requires an advanced use. In this paper we provide a simple way to fit complex spatio-temporal models for count data including sum-to-zero constraints to solve identifiability issues. Our procedure exploits some useful matrix results and the main advantage is that no extra programming is needed as we take advantage of implemented functions. The results obtained with NIMBLE and R-INLA are identical in terms of relative risk estimates and nearly identical in terms of parameter estimates. Finally, approximate Bayesian inference with R-INLA saves computing time in comparison to exact inference based on MCMC in NIMBLE.

Acknowledgements

This work has been supported by Project PID2020-113125RB-I00/ MCIN/ AEI/ 10.13039/501100011033.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Tomás Goicoa https://orcid.org/0000-0002-0588-0137
María Dolores Ugarte https://orcid.org/0000-0002-3505-8400

References

Adin, A., Martínez-Beneito, M. A., Botella-Rocamora, P., Goicoa, T., and Ugarte, M. D. (2017). Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. *Stochastic Environmental Research and Risk Assessment*, 31:403–415.

Assunção, R. M., Reis, I. A., and Oliveira, C. D. (2001). Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine*, 20(15):2319–2335.

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

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36:192–236.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174.

de Valpine, P., Paciorek, C., Ture, D., Michaud, N., Anderson-Bergma, C., Obermeyer, F., Wehrhahn Cortes, C., Rodríguez, A., Temple Lang, D., and Paganin, S. (2021). *NIMBLE User Manual*. R package manual version 0.11.1.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26:403–417.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, 94(445):247–253.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515–534.

Gelman, A., Carlin, J. B., Stern, H., and Rubin, D. B. (2013). *Bayesian Data Analysis. Third Edition*. Chapman and Hall/CRC.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.

Goicoa, T., Adin, A., Ugarte, M. D., and Hodges, J. S. (2018). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment*, 32(3):749–770.

Goicoa, T., Ugarte, M. D., Etxeberria, J., and Militino, A. F. (2012). Comparing CAR and P-spline models in spatial disease mapping. *Environmental and Ecological Statistics*, 19:573–599.

Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer New York.

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

Lawson, A. B. (2020). NIMBLE for Bayesian Disease Mapping. *Spatial and Spatio-temporal Epidemiology*, 33:100323.

Lindgren, F. and Rue, H. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

MacNab, Y. C. (2014). On identification in Bayesian disease mapping and ecological-spatial regression models. *Statistical Methods in Medical Research*, 23(2):134–155.

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

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Roberts, G. O. and Sahu, S. K. (1997). Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):291–317.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC press.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Spiegelhalter, D. J., Thomas, A., Best, N., and Way, R. (2003). *WinBUGS User Manual version 1.4*.

Ugarte, M., Adin, A., Goicoa, T., and Militino, A. (2014). On fitting spatio-temporal disease mapping models using approximate bayesian inference. *Statistical Methods in Medical Research*, 23(6):507–530.

Ugarte, M. D., Goicoa, T., Etxeberria, J., and Militino, A. F. (2012). A P-spline ANOVA type model in space-time disease mapping. *Stochastic Environmental Research and Risk Assesment*, 26(6):835–845.

Ugarte, M. D., Goicoa, T., and Militino, A. F. (2010). Spatio-temporal modelling of mortality risks using penalized splines. *Environmetrics*, 21(3–4):270–289.

Vranckx, M., Neyens, T., and Faes, C. (2021). The (in)stability of bayesian model selection criteria in disease mapping. *Spatial Statistics*, 43:100502.

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Wood, S. N., Scheipl, F., and Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, 23(3):341–360.

8 Appendix

In this section we provide NIMBLE code to fit spatio-temporal models with Type IV spatio-temporal interactions. In particular we consider an ICAR prior for the spatial effect, a RW1 for the temporal effect and the set of hyperpriors H1. NIMBLE code to fit models with Type I, II and III interactions as well as R-INLA code to fit all the models is available at https://github.com/spatialstatisticsupna/Comparing-R-INLA-and-NIMBLE.

8.1 Load and prepare the data

```
# Load NIMBLE package
library(nimble)

# Set working directory
setwd("")

# Load the data
load("BreastCancer_data.Rdata")

N <- length(unique(Data$Area))
t <- length(unique(Data$Year))

# Spatial random effects: specify adjacency and weights vectors
num <- diag(Rs)
adj <- c()

for (i in 1:dim(Rs)[1]){
  neighbours <- unname(which(Rs[i, ]==-1))
  adj <- c(adj, neighbours)
}
weights <- rep(1, length(adj))

# Temporal random effects: specify adjacency and weights vectors
weights.rw1 <- c()
adj.rw1 <- c()
num.rw1 <- c()

for(i in 1:1) {
  weights.rw1[i] <- 1; adj.rw1[i] <- i+1; num.rw1[i] <- 1
}
for(i in 2:(t-1)) {
  weights.rw1[2+(i-2)*2] <- 1; adj.rw1[2+(i-2)*2] <- i-1
  weights.rw1[3+(i-2)*2] <- 1; adj.rw1[3+(i-2)*2] <- i+1;
  num.rw1[i] <- 2
}
for(i in t:t) {
  weights.rw1[(i-2)*2 + 2] <- 1; adj.rw1[(i-2)*2 + 2] <- i-1;
  num.rw1[i] <- 1
}

# Counts and expected cases in matrix form
```

```
expected <-  matrix(Data$Expected, nrow = 50)
y <- matrix(Data$Counts, nrow = 50)
```

8.2 Nimble 1

```
############################
##  ICAR - RW1 - Type IV  ##
############################
icar.type4.h1 <- nimbleCode({
  # Priors
  alpha0 ~ dflat()
  sd.u ~ dunif(0, 100)
  tau.u <- 1/sd.u^2
  var.u <- sd.u^2
  sd.temp ~ dunif(0, 100)
  tau.temp <- 1/sd.temp^2
  var.temp <- sd.temp^2
  sd.t4 ~ dunif(0, 100)
  tau.t4 <- 1/sd.t4^2
  var.t4 <- sd.t4^2

  # Spatial random effects
  spat.u[1:N] ~ dcar_normal(adj[1:L], weights[1:L], num[1:N], 1,
                            zero_mean = 1)

  # Temporal random effects
  temp[1:t] ~ dcar_normal(adj.rw1[1:L.rw1], weights.rw1[1:L.rw1],
                          num.rw1[1:t], 1, zero_mean = 1)

  # Spatio-temporal random effects (type IV)
  for (k in 1:t){
    u.t4[1:N, k] ~ dcar_normal(adj[1:L], weights[1:L], num[1:N], 1,
                               zero_mean = 1)
  }
  for (i in 1:N){
    stType4[i, 1:t] <- u.t4[i, 1:t] %*% Achol[1:t, 1:t]
  }

  for(i in 1:N) {
    for (k in 1:t){
      y[i,k] ~ dpois(mu[i,k])
      log(mu[i,k]) <- log(E[i,k]) + alpha0 + sd.u*spat.u[i] + sd.temp*temp[k]
                      + sd.t4*stType4[i,k]
      RR[i,k] <- exp(alpha0 + sd.u*spat.u[i] + sd.temp*temp[k] + sd.t4*stType4[i,k])

      # Spatio-temporal trends
      Est[i,k] <- exp(sd.t4*stType4[i,k])

      # Compute the deviance
      Dev[i,k] <- -2*(-mu[i,k] + y[i,k]*log(mu[i,k])-lfactorial(y[i,k]))
    }
    # Spatial pattern
    Espat[i] <- exp(sd.u*spat.u[i])
  }

  # Temporal trend
  for(k in 1:t){
    Etemp[k] <- exp(sd.temp*temp[k])
    Ealpha0.temp[k] <- exp(alpha0 + sd.temp*temp[k])
  }

  # Compute the deviance
  sumDev <- sum(Dev[1:N, 1:t])
})

##################################
##  n.chains, n.iter, n.burnin  ##
##################################
```

```
num.chains <- 3
num.iter <- 30000
num.burnin <- 5000
num.thin <- 75

#####################
##  Run the model  ##
#####################

# Temporal structure matrix for a RW1 prior
Dm <- diff(diag(t),differences=1)
Rt <- t(Dm)%*%Dm

cov.Rt <- MASS::ginv(Rt)    # Generalized Inverse of Rt
chol.cov.Rt <- chol(cov.Rt)

# Define a list containing the counts
data <- list(y = y)

# Define constants
constants <- list(N = N, t = t, E = expected, L = length(adj), L.rw1 = length(adj.rw1),
                  adj = adj, num = num, weights = weights,
                  adj.rw1 = adj.rw1, num.rw1 = num.rw1, weights.rw1 = weights.rw1,
                  Achol = chol.cov.Rt, nt= N*t)

# Define initial values
inits <- list(alpha0 = rnorm(1,0,0.1), sd.u = runif(1,0,1),
              sd.temp = runif(1,0,1), sd.t4 = runif(1,0,1),
              spat.u = rnorm(N), temp = rnorm(t),
              stType4 = matrix(rnorm(N*t), nrow=N, ncol=t))

# Run the model
model <- nimbleModel(icar.type4.h1, constants = constants, data = data, inits = inits)
cmodel <- compileNimble(model)

conf <- configureMCMC(model, monitors = c('alpha0', 'sd.u', 'sd.temp', 'sd.t4',
                                          'tau.u', 'tau.temp', 'tau.t4',
                                          'var.u', 'var.temp', 'var.t4',
                                          'spat.u', 'temp', 'stType4',
                                          'Etemp', 'Espat', 'Est', 'Ealpha0.temp',
                                          'RR', 'Dev', 'sumDev', 'mu'))

MCMC <- buildMCMC(conf, enableWAIC = TRUE)
cMCMC <- compileNimble(MCMC, project=cmodel, resetFunctions = TRUE)

# Obtain the samples
typeIV.nimble1 <- runMCMC(cMCMC, niter = num.iter, nburnin = num.burnin,
                          nchains = num.chains, thin = num.thin,
                          samplesAsCodaMCMC = TRUE, summary=TRUE, WAIC = TRUE)
```

8.3 Nimble 2

```
############################
##  ICAR - RW1 - Type IV  ##
############################
icar.type4.h1 <- nimbleCode({
  # Priors
  alpha0 ~ dflat()
  sd.u ~ dunif(0, 100)
  tau.u <- 1/sd.u^2
  var.u <- sd.u^2
  sd.temp ~ dunif(0, 100)
  tau.temp <- 1/sd.temp^2
  var.temp <- sd.temp^2
  sd.t4 ~ dunif(0, 100)
  tau.t4 <- 1/sd.t4^2
  var.t4 <- sd.t4^2
```

```
  # Spatial random effects
  spat.u[1:N] ~ dcar_normal(adj[1:L], weights[1:L], num[1:N], 1, zero_mean = 1)

  # Temporal random effects
  temp[1]  <- temp.zero
  for (k in 2:t){
    temp[k] ~ dnorm(temp[k-1], 1)
  }

  # Spatio-temporal random effects (type IV)
  for (k in 1:t){
    u.t4[1:N, k] ~ dcar_normal(adj[1:L], weights[1:L], num[1:N], 1, zero_mean = 1)
  }
  for (i in 1:N){
    stType4[i, 1:t] <- u.t4[i, 1:t] %*% Achol[1:t, 1:t]
  }

  for(i in 1:N) {
    for (k in 1:t){
      y[i,k] ~ dpois(mu[i,k])
      log(mu[i,k]) <- log(E[i,k]) + alpha0 + sd.u*spat.u[i] + sd.temp*temp[k]
                      + sd.t4*stType4[i,k]
      RR[i,k] <- exp(alpha0 + sd.u*spat.u[i] + sd.temp*temp[k] + sd.t4*stType4[i,k])

      # Spatio-temporal trend
      Est[i,k] <- exp(sd.t4*stType4[i,k])

      # Compute the deviance
      Dev[i,k] <- -2*(-mu[i,k] + y[i,k]*log(mu[i,k])-lfactorial(y[i,k]))
    }
    # Spatial pattern
    Espat[i] <- exp(sd.u*spat.u[i])
  }

  # Temporal trend
  for(k in 1:t){
    Etemp[k] <- exp(sd.temp*temp[k])
    Ealpha0.temp[k] <- exp(alpha0 + sd.temp*temp[k])
  }

  # Compute the deviance
  sumDev <- sum(Dev[1:N, 1:t])
})


###################################
##  n.chains, n.iter, n.burnin  ##
###################################
num.chains <- 3
num.iter <- 30000
num.burnin <- 5000
num.thin <- 75

#####################
##  Run the model  ##
#####################

# Temporal structure matrix for a RW1 prior
Dm <- diff(diag(t),differences=1)
Rt <- t(Dm)%*%Dm

cov.Rt <- MASS::ginv(Rt)    # Generalized Inverse of Rt
chol.cov.Rt <- chol(cov.Rt)

# Define a list containing the counts
data <- list(y = y)
```

```
# Define constants
constants <- list(N = N, t = t, E = expected, L = length(adj),
                  adj = adj, num = num, weights = weights,
                  Achol = chol.cov.Rt, nt= N*t, temp.zero = 0)

# Define initial values
inits <- list(alpha0 = rnorm(1,0,0.1), sd.u = runif(1,0,1),
              sd.temp = runif(1,0,1), sd.t4 = runif(1,0,1),
              spat.u = rnorm(50), temp = c(0, rnorm(20)),
              stType4 = matrix(rnorm(N*t), nrow=N, ncol=t))

# Run the model
model <- nimbleModel(icar.type4.h1, constants = constants, data = data, inits = inits)
cmodel <- compileNimble(model)

conf <- configureMCMC(model, monitors = c('alpha0', 'sd.u', 'sd.temp', 'sd.t4',
                                          'tau.u', 'tau.temp', 'tau.t4',
                                          'var.u', 'var.temp', 'var.t4',
                                          'spat.u', 'temp', 'stType4',
                                          'Etemp', 'Espat', 'Est', 'Ealpha0.temp',
                                          'RR', 'Dev', 'sumDev', 'mu'))

MCMC <- buildMCMC(conf, enableWAIC = TRUE)
cMCMC <- compileNimble(MCMC, project=cmodel, resetFunctions = TRUE)

# Obtain the samples
typeIV.nimble2 <- runMCMC(cMCMC, niter = num.iter, nburnin = num.burnin,
                          nchains = num.chains, thin = num.thin,
                          samplesAsCodaMCMC = TRUE, summary=TRUE, WAIC = TRUE)
```

8.4 Compute the DIC

```
DIC.nimble <- function(summary){
  dev <- summary[grep('sumDev', rownames(summary), value=TRUE), ]
  mu <- summary[grep('mu', rownames(summary), value=TRUE), ]

  dev2 <- 0
  for (i in 1:(N*t)){
    dev2 <- dev2 -2*(-mu[i, 1] + Data$Counts[i]*log(mu[i, 1])
             - lfactorial(Data$Counts[i]))
  }
  pD <- dev[1] - dev2
  data.frame(mean.dev=dev[1],      # Mean deviance
             pD=dev[1] - dev2,     # Effective number of parameters
             DIC=dev[1]+pD)
}

DIC.nimble(typeIV.nimble1$summary$all.chains)
DIC.nimble(typeIV.nimble2$summary$all.chains)
```

Supplementary material for

"Space-time interactions in Bayesian disease mapping with recent tools: making things easier for practitioners"

by Arantxa Urdangarin, Tomás Goicoa, María Dolores Ugarte

A Introduction

This supplementary material contains tables and figures to complement the paper entitled "Space-time interactions in Bayesian disease mapping with recent tools: making things easier for practitioners".

Table A.1: Mean deviance ($\overline{D(\theta)}$), effective number of parameters (p_D), DIC and WAIC for spatio-temporal models with an ICAR and a BYM spatial prior and the set of hyperpriors **H2**.

		ICAR spatial prior				**BYM spatial prior**			
		$\overline{D(\theta)}$	p_D	DIC	$WAIC$	$\overline{D(\theta)}$	p_D	DIC	$WAIC$
Type I	R-INLA	7603.4820	214.0369	7817.5190	7825.0270	7602.2280	215.9365	7818.1650	7824.2920
	Nimble 1	7605.1050	215.6115	7820.7170	7829.7650	7606.4940	213.1343	7819.6280	7828.1740
	Nimble 2	7608.1640	210.4046	7818.5690	7828.5620	7604.4650	212.3013	7816.7660	7825.1430
Type II	R-INLA	7551.0570	160.9670	7712.0240	7714.4540	7550.0980	161.2665	7711.3650	7713.7950
	Nimble 1	7554.1190	161.9542	7716.0730	7720.6500	7551.7440	162.6510	7714.3950	7718.0410
	Nimble 2	7550.8400	164.6176	7715.4580	7718.4830	7552.3060	160.7408	7713.0470	7716.0380
Type III	R-INLA	7608.3230	182.4803	7790.8040	7801.3280	7607.3880	183.5559	7790.9440	7800.6290
	Nimble 1	7611.9600	181.5406	7793.5010	7805.8850	7611.7120	181.3514	7793.0640	7804.1160
	Nimble 2	7603.8980	189.1573	7793.0550	7803.9930	7610.9670	183.7092	7794.6760	7807.4970
Type IV	R-INLA	7560.3550	143.6618	7704.0170	7706.7050	7559.2620	144.4739	7703.7360	7705.9290
	Nimble 1	7562.8260	144.1419	7706.9680	7710.7630	7561.3760	143.7267	7705.1020	7708.7370
	Nimble 2	7557.6240	149.8634	7707.4880	7711.3580	7560.6600	144.3094	7704.9690	7708.2910

Table A.2: Mean deviance ($\overline{D(\theta)}$), effective number of parameters (p_D), DIC and WAIC for spatio-temporal models with an ICAR and a BYM spatial prior and the set of hyperpriors **H3**.

		ICAR spatial prior				**BYM spatial prior**			
		$\overline{D(\theta)}$	p_D	DIC	$WAIC$	$\overline{D(\theta)}$	p_D	DIC	$WAIC$
Type I	R-INLA	7603.3559	214.1094	7817.4653	7824.9643	7601.8410	214.7837	7816.6250	7824.0930
	Nimble 1	7607.7301	212.1685	7819.8986	7830.2154	7608.8370	210.6310	7819.4680	7828.5590
	Nimble 2	7607.0330	211.5730	7818.6060	7828.5590	7607.1920	212.4161	7819.6090	7828.5040
Type II	R-INLA	7550.9084	161.0542	7711.9626	7714.3817	7549.1380	161.9340	7711.0720	7713.4560
	Nimble 1	7553.3957	161.9452	7715.3409	7718.4217	7552.8830	160.5058	7713.3890	7717.0800
	Nimble 2	7551.7970	158.2855	7710.0830	7712.9970	7552.0320	161.0963	7713.1280	7716.1560
Type III	R-INLA	7608.2090	182.5368	7790.7458	7801.2679	7606.4920	183.4056	7789.8970	7800.4370
	Nimble 1	7611.9792	184.0053	7795.9845	7808.3194	7609.0680	181.9525	7791.0200	7803.1340
	Nimble 2	7608.6500	182.5053	7791.1560	7803.6660	7608.9170	181.8855	7790.8030	7803.4300
Type IV	R-INLA	7560.2337	143.7245	7703.9582	7706.6419	7558.4520	144.6587	7703.1110	7705.8110
	Nimble 1	7563.5695	144.5272	7708.0967	7712.1708	7561.6870	144.0091	7705.6960	7709.2910
	Nimble 2	7563.5600	143.3191	7706.8790	7710.8980	7562.0640	144.67530	7706.7400	7711.0160

Table A.3: Posterior means and standard deviations of the intercept and the hyperparameters for the model with the BYM spatial prior, Type IV interaction and the three sets of hyperprior distributions.

		H1		**H2**		**H3**	
		Mean	SD	Mean	SD	Mean	SD
α_0	R-INLA	-0.0353	0.0106	-0.0349	0.0096	-0.0353	0.0116
	Nimble 1	-0.0344	0.0102	-0.0350	0.0082	-0.0354	0.0136
	Nimble 2	0.0566	0.0152	0.0567	0.0135	0.0553	0.0167
σ_u	R-INLA	0.1454	0.0360	0.1436	0.0276	0.1206	0.0289
	Nimble 1	0.1483	0.0404	0.1577	0.0530	0.1236	0.0415
	Nimble 2	0.1543	0.0378	0.1840	0.0310	0.1226	0.0305
σ_v	R-INLA	0.0686	0.0173	0.0573	0.0135	0.0779	0.0150
	Nimble 1	0.0591	0.0255	0.0359	0.0335	0.0792	0.0240
	Nimble 2	0.0559	0.0250	0.0202	0.0234	0.0762	0.0143
σ_γ	R-INLA	0.0246	0.0052	0.0217	0.0040	0.0219	0.0044
	Nimble 1	0.0240	0.0052	0.0221	0.0047	0.0222	0.0045
	Nimble 2	0.0252	0.0063	0.0219	0.0042	0.0220	0.0045
σ_δ	R-INLA	0.0389	0.0046	0.0373	0.0045	0.0374	0.0045
	Nimble 1	0.0389	0.0046	0.0373	0.0043	0.0370	0.0045
	Nimble 2	0.0388	0.0047	0.0379	0.0043	0.0372	0.0045

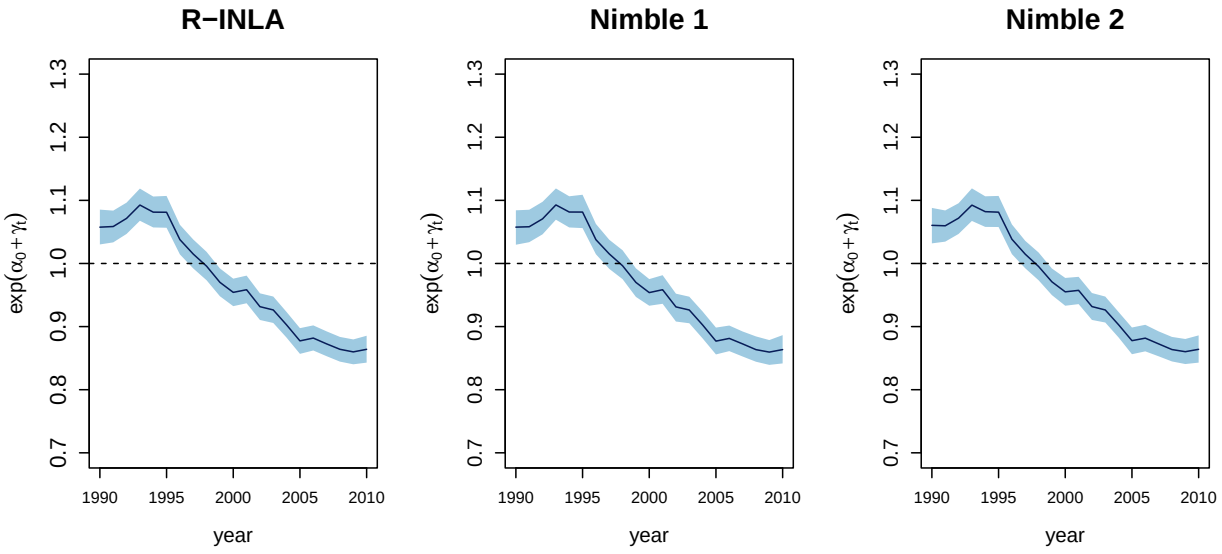

Figure A.1: Posterior means of $\exp(\alpha_0 + \gamma_t)$ estimated with R-INLA, Nimble 1 and Nimble 3 for ICAR model with H1 hyperpriors.

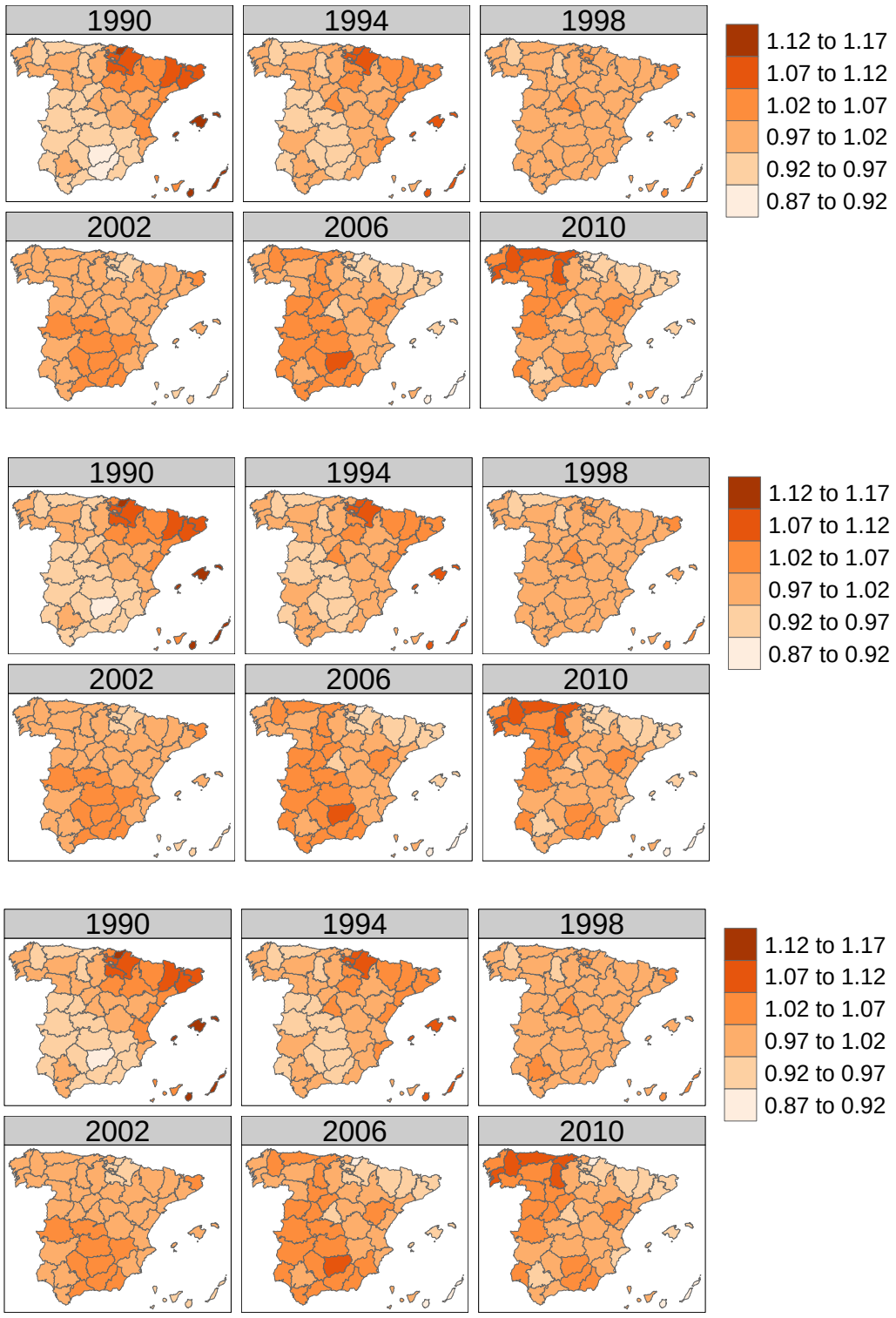

Figure A.2: Spatio-temporal patterns (posterior means of $\exp(\delta_{it})$) estimated with R-INLA (top), Nimble 1 (middle), and Nimble 2 (bottom) for ICAR model with H1 hyperpriors.

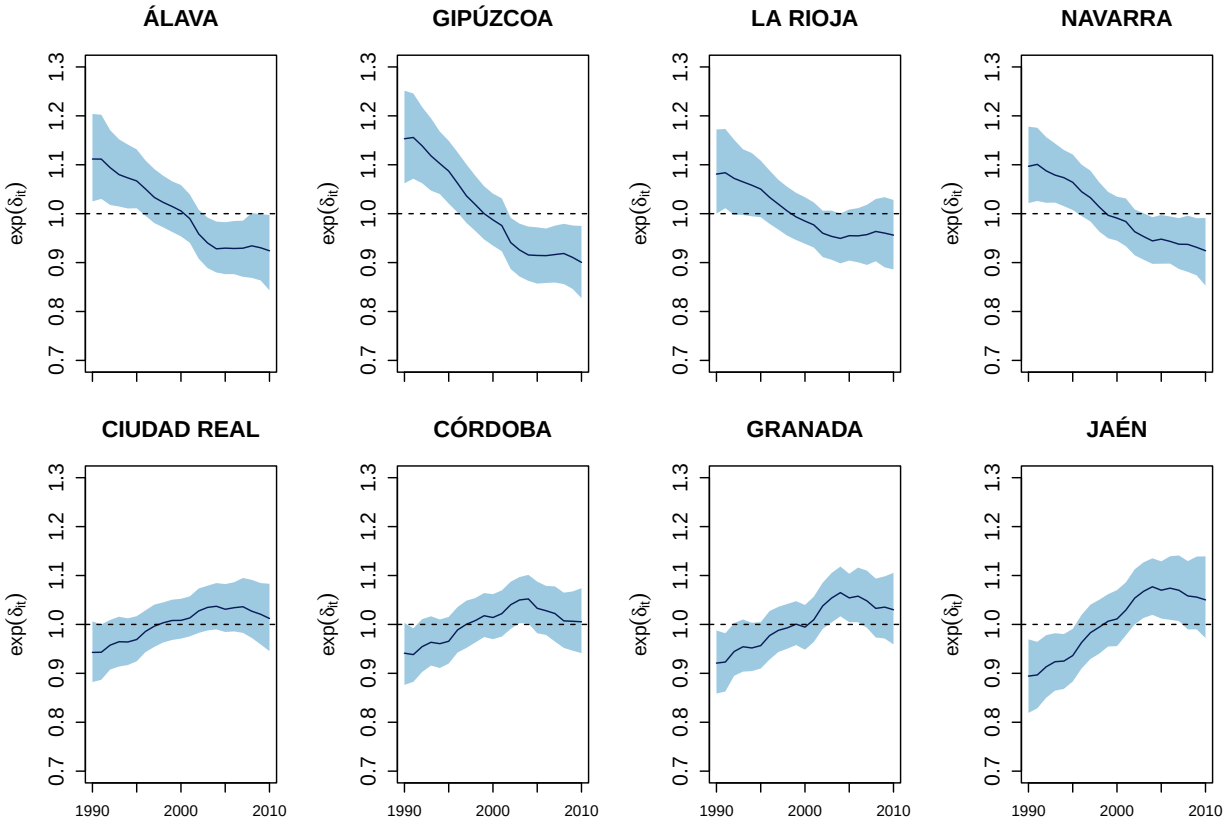

Figure A.3: Type IV space-time interaction random effects (posterior means of $\exp(\delta_{it})$) estimated with Nimble 1 and their credibility bands. Top row corresponds to neighbouring provinces in the north and bottom row displays neighbouring provinces in the south.

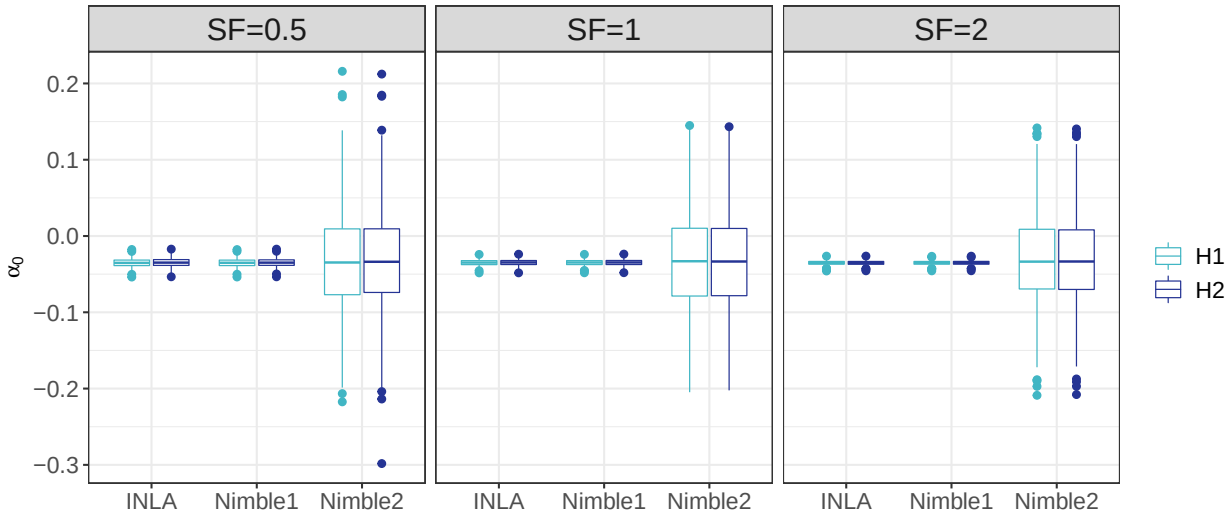

Figure A.4: Boxplots of the posterior means of the intercept in the 500 simulations.