Parallel and distributed Bayesian modelling for analysing high-dimensional spatio-temporal count data

Orozco-Acosta, E\textsuperscript{1,2}, Adin, A\textsuperscript{1,2} and Ugarte, M.D.\textsuperscript{1,2}

\textsuperscript{1} Department of Statistics, Computer Sciences and Mathematics, Public University of Navarre, Spain.
\textsuperscript{2} Institute for Advanced Materials and Mathematics, InaMat\textsuperscript{2}, Public University of Navarre, Spain.

∗Correspondence to María Dolores Ugarte, Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra, Campus de Arrosadia, 31006 Pamplona, Spain.

\textbf{E-mail:} lola@unavarra.es

Abstract

This paper proposes a general procedure to analyse high-dimensional spatio-temporal count data, with special emphasis on relative risks estimation in cancer epidemiology. More precisely, we present a pragmatic and simple idea that permits to fit hierarchical spatio-temporal models when the number of small areas is very large. Model fitting is carried out using integrated nested Laplace approximations over a partition of the spatial domain. We also use parallel and distributed strategies to speed up computations in a setting where Bayesian model fitting is generally prohibitively time-consuming and even unfeasible. The whole procedure is evaluated in a simulation study with a twofold objective: to estimate risks accurately and to detect extreme risk areas while avoiding false positives/negatives. We show that our method outperforms classical global models. A real data analysis comparing the global models and the new procedure is also presented.

Keywords: Disease mapping; Hierarchical models; Laplace approximations; Massive data; Non-stationary models; Scalable modelling;

1 Introduction

In recent decades, access to geospatial data through Geographical Information Systems (GIS) and other related technologies has grown at a staggering rate. Modern geospatial data typically involve large datasets collected from a variety of sources (databases or servers) that may include information such as satellite imagery, weather data, census data, social network data and public health data. Consequently, the development of new techniques and computational algorithms to analyse massive spatial and spatio-temporal datasets is of crucial interest in many fields such as remote sensing, geoscience, ecology, crime research and epidemiology among others.

Hierarchical spatial models including random effects (Cressie and Wikle, 2011; Banerjee et al., 2015) are widely used in spatial statistics to provide reliable estimates of the underlying geographical phenomenon and quantifying uncertainty in predictions at unobserved locations. See, for example, Sun et al. (2012) and Banerjee (2017) for a detailed review of methods and scalable models for high-dimensional spatial and spatio-temporal data. Gaussian processes (GPs)
have been commonly used for the analysis of geostatistical (point-referenced) data in the spatial statistics literature. However, traditional estimation of GPs has become computationally intractable when analysing modern big datasets, mainly due to computations involving matrix factorizations for very large covariance matrices. During the last years, many approaches have been proposed to ensure scalability of large geostatistical datasets (see, e.g., Heaton et al. (2019) and Liu et al. (2020) for recent reviews and comparisons). Some other recent methods to deal with massive datasets are described below. Appel and Pebesma (2020) provide an extension to the multi-resolution approximation approach (Katzfuss, 2017) for spatio-temporal modelling of global datasets, where a recursive partitioning scheme is considered so that inference can be efficiently scaled in distributed computing environments. Zammit-Mangion and Rougier (2020) propose an approximate inference scalable algorithm for multi-scale process modelling by using the stochastic partial differential equation approach (Lindgren et al., 2011). Both methods were applied to modelling and prediction of global sea-surface temperature. In the geostatistical literature, many recent works are being proposed to estimate GPs based on the so-called Vecchia approximation. (Vecchia, 1988). This approximation can be regarded as a special case of the Gaussian Markov random field approximations (Rue and Held, 2005) with a simplified neighbourhood structure that can be represented by directed acyclic graph (DAG) models. This representation leads to a sparse formulation of the precision matrix which ensures that evaluating the likelihood of the GPs will be computationally scalable. Based on this approach, Finley et al. (2019) propose alternative formulations of Bayesian nearest neighbour Gaussian process models developed by Datta et al. (2016) to substantially improve computational efficiency; Peruzzi et al. (2020) develop a meshed Gaussian process with a novel partitioning and graph design based on domain tessellations while Katzfuss and Guinness (2021) propose a novel sparse general Vecchia approximation algorithm which ensures computational feasibility for large spatial datasets; Jurek and Katzfuss (2022) present a fast and simple algorithm to compute their hierarchical Vecchia approximation, and provide extensions to nonlinear data assimilation with non-Gaussian data based on the Laplace approximation.

Although there is an extensive literature developing scalable methods and computational algorithms for analysing massive geostatistical data, only a few papers discuss scalable disease mapping models for high dimensional areal data. Disease mapping is the field of spatial epidemiology that deals with aggregated count data from non-overlapping areal units focussing on the estimation of the geographical distribution of a disease and its evolution in time (Lawson et al., 2016). As outlined by Shen and Louis (2000), the three main inferential goals in disease mapping are: (i) to provide accurate estimates of mortality/incidence risks or rates in space and time, (ii) to unveil the underlying spatial and spatio-temporal patterns, and (iii) to detect high-risk areas or hotspots. Since classical risk estimation measures, such as the standardized mortality/incidence ratio, are extremely variable when analysing rare diseases (with very few cases) or low-populated areas, several statistical models have been proposed during the last decades to obtain smooth disease risk estimates borrowing information from spatial and/or temporal neighbours. Research into spatial and spatio-temporal disease mapping has been carried out within a hierarchical Bayesian framework, with generalized linear mixed models (GLMMs) playing a major role. Although GLMMs including spatial and temporal random effects are a very popular and flexible approach to model areal count data, these smoothing methods become computationally challenging (or even unfeasible) when analysing very large spatio-temporal datasets. Guan and Haran (2018) develop a method to reduce the dimension of the spatial random effect by reparameterizing the model based on random projections of the covariance matrix. In
addition, they show how to address confounding issues if explanatory variables are included in the model by simultaneously applying the restricted spatial regression approach (Hodges and Reich, 2010). This model is similar to the one proposed by Hughes and Haran (2013), where the decomposition is performed based on the Moran operator. Datta et al. (2019) introduce a class of directed acyclic graphical autoregression (DAGAR) models as an alternative to the commonly used conditional autoregressive (CAR) models for spatial areal data. Instead of modelling the precision matrix of the spatial random effect, they propose to model its (sparse) Cholesky factor using autoregressive covariance models on a sequence of local trees created from the directed acyclic graph derived from the original undirected graph (spatial neighbourhood structure) of the areal units. As stated by the authors, the Cholesky factor has the same level of sparsity as the undirected graph ensuring scalability for analysing very large areal datasets. An extension to deal with multivariate spatial disease mapping models has been developed by Gao et al. (2021).

Very recently, a scalable Bayesian spatial model has been proposed by Orozco-Acosta et al. (2021) based on the “divide-and-conquer” approach so that local spatial CAR models can be simultaneously fitted. This new methodology provides reliable risk estimates with a substantial reduction in computational time.

The modelling approaches described above are limited to the analysis of spatial count data. The main objective of this paper is to propose a scalable Bayesian modelling approach to smooth mortality or incidence risks in in a high-dimensional spatio-temporal disease mapping context by extending the methodology described in Orozco-Acosta et al. (2021). Specifically, we adapt the modelling scheme so that commonly used spatio-temporal models can be fitted over different subdomains (partitions of the region of interest), which allows to define non-stationary models, i.e., models that induce different degree of smoothing over the areal units belonging to each subdomain. From a theoretical point of view, both spatial and/or temporal partitions of the data could be defined, however, in the disease mapping context the high-dimensionality of the data is usually related to the estimation of relative risks at a fine-scale spatial resolution. The main challenges of the methodology presented in this work is not only to extend the “divide-and-conquer” approach to deal with spatio-temporal models (which is not trivial at all), but also to derive and implement specific algorithms to perform scalable model estimation in both parallel or distributed processing architectures.

The remainder of this article is organized as follows. Section 2 poses the spatio-temporal CAR models considered in this work. Section 3 introduces the new scalable Bayesian models and describes a generic scheme of the main algorithms that have been implemented in this work. In Section 4 a simulation study based on a template of over the almost 8000 municipalities of continental Spain and 25 time periods is conducted to compare the new scalable methods with previous proposals. In addition, a numerical simulation is conducted to evaluate the computational gain offered by our modelling approach when the number of small areas increases. In Section 5 we use the new model proposal to analyse lung cancer mortality data in Spanish municipalities. The paper ends with a discussion.

## 2 Spatio temporal models in disease mapping

Let us assume that the region under study is divided into contiguous small areas labelled as \( i = 1, \ldots, n \) and data are available for consecutive time periods labelled as \( t = 1, \ldots, T \). For a given area \( i \) and time period \( t \), \( O_{it} \) and \( E_{it} \) denote the number of observed and expected cases,
respectively. To compute the number of expected cases both direct and indirect standardization methods can be used, usually considering age and/or sex as standardization variables. When using the indirect method, the number of expected cases for area $i$ and time $t$ is calculated as

$$E_{it} = \sum_{j=1}^{J} N_{itj} O_{j} / N_{j}$$

for $i = 1, \ldots, n; t = 1, \ldots, T$,

where $O_{j} = \sum_{i=1}^{n} \sum_{t=1}^{T} O_{itj}$ and $N_{j} = \sum_{i=1}^{n} \sum_{t=1}^{T} N_{itj}$ are the number of observed cases and the population at risk in the $j$th age-and-sex group, respectively. Then, the standardized mortality/incidence ratio (SMR or SIR) is defined as the number of observed cases divided by the number of expected cases. Although its interpretation is very simple, SMRs are extremely variable when analysing rare diseases or very low-populated areas, as it is the case of high-dimensional data. This makes it necessary the use of statistical models to smooth risks borrowing information from neighbouring regions and time periods.

Poisson mixed models are typically used for the analysis of count data within a hierarchical Bayesian framework. Conditional to the relative risk $r_{it}$, the number of observed cases in the $i$th area and time period $t$ is assumed to be Poisson distributed with mean $\mu_{it} = E_{it} r_{it}$. That is,

$$O_{it} | r_{it} \sim \text{Poisson}(\mu_{it} = E_{it} r_{it}),$$

$$\log \mu_{it} = \log E_{it} + \log r_{it},$$

where $\log E_{it}$ is an offset. Depending on the specification of the log-risks different models can be defined.

Probably, the non-parametric models based on CAR priors for spatial random effects, random walk priors for temporal random effects, and different types of spatio-temporal interactions described in [Knorr-Held (2000)] are the most widely used models in space-time disease mapping. Slight modifications of these models are considered here, so the log-risks are modelled as

$$\log r_{it} = \alpha + \xi_{i} + \gamma_{t} + \delta_{it},$$

(1)

where $\alpha$ is an intercept representing the overall log-risk, $\xi_{i}$ is a spatial random effect with CAR prior distribution, $\gamma_{t}$ is a temporally structured random effect that follows a random walk prior distribution, and $\delta_{it}$ is a spatio-temporal random effect. All the components of this model can be modelled as GMRFs and prior densities can be written according to some structure matrices.

A modification of the [Dean et al. (2001)] model proposed by [Riebler et al. (2016)], hereafter called BYM2 model, has been considered as the prior distribution for the spatial random effects $\xi = (\xi_{1}, \ldots, \xi_{n})^{t}$, so that

$$\xi \sim N(0, Q_{\xi}^{*})$$

with $Q_{\xi}^{*} = \tau_{\xi}^{-1}[(1 - \lambda_{\xi})I_{n} + \lambda_{\xi}R_{\xi}]$,
where $I_n$ is the $n \times n$ identity matrix, and $R_\gamma^-$ indicates the generalised inverse of the scaled spatial structure matrix corresponding to the undirected graph of the regions under study (see, e.g., Sorbye and Rue [2014]). Recall that the spatial structure matrix is defined as $R_\xi = D_W - W$, where $D_W = \text{diag}(w_1, \ldots, w_n)$ and $w_1 = \sum_j w_{ij}$ is the $i^{th}$ row sum of the binary adjacency matrix $W = (w_{ij})$, whose $ij^{th}$ element is equal to one if areas $i$ and $j$ are defined as neighbours (usually if they share a common border), and it is zero otherwise.

For the temporally structured random effect $\gamma = (\gamma_1, \ldots, \gamma_T)'$, random walks of first (RW1) or second order (RW2) prior distributions can be assumed as follow

$$\gamma \sim N(0, [\tau_\gamma R_\gamma]^-),$$

where $\tau_\gamma$ is a precision parameter and $R_\gamma$ is the $T \times T$ structure matrix of a RW1/RW2 (see Rue and Held [2005], pp. 95 and 110).

Finally, for the space-time interaction random effect $\delta = (\delta_{11}, \ldots, \delta_{n1}, \ldots, \delta_{1T}, \ldots, \delta_{nT})'$ the following prior distribution is assumed

$$\delta \sim N(0, [\tau_\delta R_\delta]^-),$$

where $\tau_\delta$ is a precision parameter and $R_\delta$ is the $nT \times nT$ matrix obtained as the Kronecker product of the corresponding spatial and temporal structure matrices, where four different types of interactions where originally proposed by Knorr-Held (2000) (see Table 1).

Table 1: Specification for different types of space-time interactions.

| Interaction | $R_\delta$ | Spatial correlation | Temporal correlation |
|-------------|------------|---------------------|---------------------|
| Type I      | $I_T \otimes I_n$ | –                    | –               |
| Type II     | $R_\gamma \otimes I_n$ | – | ✓ |
| Type III    | $I_T \otimes R_\xi$ | ✓ | – |
| Type IV     | $R_\gamma \otimes R_\xi$ | ✓ | ✓ |

In what follows, we will refer to Model (1) as the Global model. These models are flexible enough to describe many real situations, and their interpretation is simple and attractive. However, the models are typically not identifiable and appropriate sum-to-zero constraints must be imposed over the random effects [Goicoa et al., 2018]. See Table 2 for a full description of the identifiability constraints that need to be imposed on each type of space-time interaction.

2.1 Model fitting via integrated nested Laplace approximations

Bayesian inference has traditionally been used to fit spatial and spatio-temporal disease mapping models. The fully Bayesian approach provides posterior distributions of model parameters instead of a single point estimate. However, these distributions cannot usually be derived analytically and simulation techniques based on Markov chain Monte Carlo (MCMC) methods have been traditionally used for Bayesian inference [Gilks et al., 1995]. Although these simulation-based techniques are widely used, mainly due to the development of free software to run MCMC
Table 2: Identifiability constraints for the different types of space-time interaction effects in CAR models (Goicoa et al., 2018).

| Interaction | $R_δ$ | Constraints |
|-------------|-------|-------------|
| Type I      | $I_T ⊗ I_n$ | $\sum_{i=1}^{S} \xi_i = 0$, $\sum_{t=1}^{T} \gamma_t = 0$, and $\sum_{i=1}^{S} \sum_{t=1}^{T} \delta_{it} = 0$ |
| Type II     | $R_γ ⊗ I_n$ | $\sum_{i=1}^{S} \xi_i = 0$, $\sum_{t=1}^{T} \gamma_t = 0$, and $\sum_{i=1}^{S} \sum_{t=1}^{T} \delta_{it} = 0$, for $i = 1, \ldots, S$ |
| Type III    | $I_T ⊗ R_\xi$ | $\sum_{i=1}^{S} \xi_i = 0$, $\sum_{t=1}^{T} \gamma_t = 0$, and $\sum_{i=1}^{S} \sum_{t=1}^{T} \delta_{it} = 0$, for $t = 1, \ldots, T$ |
| Type IV     | $R_γ ⊗ R_\xi$ | $\sum_{i=1}^{S} \xi_i = 0$, $\sum_{t=1}^{T} \gamma_t = 0$, and $\sum_{i=1}^{S} \sum_{t=1}^{T} \delta_{it} = 0$, for $i = 1, \ldots, S$, $\sum_{i=1}^{S} \sum_{t=1}^{T} \delta_{it} = 0$, for $t = 1, \ldots, T$. |

3 Scalable Bayesian model proposal

There is no doubt that the use of spatio-temporal CAR models allows to obtain accurate risk estimates in reasonable computational times when the number of areal-units is relatively small. However, two main issues arise when analyzing very large spatio-temporal datasets: (i) computational time and resources, and (ii) model assumptions. Most of the smoothing methods proposed in the literature (including CAR models) are built on the idea of spatial/temporal correlation and generally use a covariance or precision matrix with dimension equal to the number of data (spatial locations × time points), leading to prohibitive computational times if (partial) matrix inversions are necessary during the estimation process. In addition, CAR models induce the same degree of spatial dependence through the whole adjacency graph. However, the larger a spatial domain is, the less likely is that the data are stationary across the whole map.

With the objective of overcoming these problematic aspects, we propose a scalable and non-
stationary Bayesian modelling approach by extending the models described in Orozco-Acosta et al. (2021) based on the idea of “divide-and-conquer” so that local spatio-temporal models can be simultaneously fitted. Our modelling approach consists of three main steps. First, the region of interest is divided into \( D \) subdomains. Then, local spatio-temporal models are fitted using a fully Bayesian approach based on INLA. Finally, the results are merged to obtain posterior marginal estimates of the relative risks for each areal-time unit. Instead of considering global random effects whose correlation structures are based on the whole spatial/temporal neighbourhood graphs of the areal-units, as is the case of the Global model described in Equation (1), we propose to divide the data into subdomains based on spatial partitions so that models with different local correlation structures, that is, models inducing different amount of smoothing are defined. Then, extending the methodology described in Orozco-Acosta et al. (2021), Disjoint and \( k \)-order neighbourhood models are defined for estimating spatio-temporal disease risks.

For the Disjoint model, a partition of the spatial domain \( D \) into \( D \) subdomains is defined, so that \( D = \bigcup_{d=1}^{D} D_d \) where \( D_j \cap D_k = \emptyset \) for all \( j \neq k \). If we denote as \( A_t \) to the small area \( i \) in time period \( t \), let \( O_d = \{O_{it} | A_{it} \in D_d\} \) and \( E_d = \{E_{it} | A_{it} \in D_d\} \) represent the observed and expected number of disease cases in each subdomain, respectively. Then, \( D \) independent local spatio-temporal models similar to those described in Section 2 are simultaneously fitted. Since each areal-time unit \( A_{it} \) belongs to a single subdomain, the final log-risk surface \( \log r = (\log r_1, \ldots, \log r_D)' \) is just the union of the posterior marginal estimates of each spatio-temporal sub-model.

However, assuming independence between areal-time units belonging to different subdomains could be very restrictive and may lead to border effects. To avoid this undesirable issue, we define the \( k \)-order neighbourhood model by adding neighbouring areal units (based on spatial adjacency) to each partition. A toy example of a spatial partition into four subdomains is represented in Figure 1. Since multiple log-risk estimates can be obtained for some \( A_{it} \) units from the different local submodels, their posterior estimates must be properly combined to obtain a single posterior distribution for each \( r_{it} \). Originally, Orozco-Acosta et al. (2021) propose to compute mixture distributions of the estimated posterior probability density functions with weights proportional to the conditional predictive ordinates (CPOs; Pettit, 1990). Here, we also investigate the strategy of using the posterior marginal risk estimates of \( A_{it} \) corresponding to the original domain the \( i \)-th area belonged to. A full comparison in terms of risk estimation accuracy and high/low risk area detection using these two merging strategies is described in Section 4.

3.1 Model implementation in the \( \text{R} \) package bigDM

Several scalable spatial and spatio-temporal models for high-dimensional areal count data are implemented in the \( \text{R} \) package bigDM (Adin et al., 2022). A generic scheme of the main algorithms implemented in the bigDM package to fit scalable spatio-temporal disease mapping models are described in Algorithms 1, 2 and 3.

Since in the disease mapping context the high-dimensionality of the data is usually related to a large number of small areas, only purely spatial partitions are implemented in the bigDM package. These partitions could be based on administrative divisions of the area of interest (such as provinces, states or local health areas), or random partitions based on a regular grid over the associated cartography. However, random partitions should be carefully done, since small domains with large number of areas with no observed cases could lead to wrong model
When fitting both the Disjoint and k-order neighbourhood models, parallel or distributed computation strategies can be performed to speed up computations by using the future package (Bengtsson, 2021). If the plan="sequential" argument is specified, the models are fitted one at a time in the current R session (local machine). In contrast, if the plan="cluster" argument is defined, multiple models can be fitted in parallel on external R sessions (local machine) or distributed in remote computing nodes. When using this option, the identifications of the local/remote workers where the models are going to be processed must be configured through the workers argument. As it is well known, the communication between the “master node” and the rest of workers affects the computational time, so the decision on how to configure the processing architecture must be made carefully (depending on the characteristics of the computations to be performed).

As described in the previous section, two different merging strategies could be considered to properly combined the posterior marginal estimates of the relative risks when fitting the k-order neighbourhood models. If the merge.strategy="mixture" argument is specified, mixture distributions of the posterior probability density functions are computed for each log$r_{it}$ (see Algorithms 2). On the other hand, if the merge.strategy="original" argument is specified, the posterior marginal estimate of the areal-unit corresponding to the original subdomain is selected.

In addition, approximations to model selection criteria such as the deviance information criterion (DIC) (Spiegelhalter et al. 2002) and the Watanabe-Akaike information criterion (WAIC) (Watanabe 2010), two widely used criteria to compare models in a fully Bayesian setting, are also derived by default when fitting the scalable models using the bigDM package. For further details and examples please see the vignettes accompanying the package available at https://github.com/spatialstatisticsupna/bigDM
Algorithm 1 Fit a scalable spatio-temporal model for high-dimensional areal count data.

**Inputs:**
- Cartography file with count data corresponding to areal units $A_{it}$, for $i = 1, \ldots, n$, and $t = 1, \ldots, T$.
- Observed cases $O_{it}$ and expected cases $E_{it}$.
- Prior distributions for the spatial ($\xi$), temporal ($\gamma$) and spatio-temporal ($\delta$) random effects.
- $W$: binary adjacency matrix of the spatial areal units.
- $k$: numeric value with the neighbourhood order.
- **plan**: computation strategy used for model fitting (one of either "sequential" or "cluster").
- **workers**: IDs of the local or remote workers (only required if plan="cluster").
- **merge.strategy**: merging strategy to compute posterior marginal estimates of relative risks. One of either "mixture" or "original" (default).

**Step 1: Pre-processing the data**
1: if $W$=NULL then
2:  compute $W$ from the cartography file.
3:  Merge disjoint connected subgraphs.
4: Define formula object for INLA model according to the prior distributions for $\xi$, $\gamma$ and $\delta$.

**Step 2: Fitting submodels with INLA**
1: Divide the spatial domain into $D$ subdomains.
2: for $d \in \{1, \ldots, D\}$ do
3:  if $k>0$ then
4:     add $k$-order neighbouring areas.
5:  Compute the spatial adjacency matrix $W_d$.
6:  Extract $O_d = \{O_{it}|A_{it} \in D_d\}$ and $E_d = \{E_{it}|A_{it} \in D_d\}$.
7:  Define appropriate identifiability constraints.
8:  if **plan="sequential"** then
9:      fit INLA models sequentially in the current R session (local machine).
10: else
11:      fit INLA models in parallel on external R sessions (local machine) or distributed in remote compute nodes.

**Step 3: Merging results**
1: if **plan="cluster"** then
2:  retrieve submodels to the central node.
3:  if $k>0$ and **merge.strategy="mixture"** then
4:       compute mixture distributions for the posterior probability density functions of each $\log r_{it}$.
\[ \Rightarrow \text{Algorithm 2} \]
5:  if $k>0$ and **merge.strategy="original"** then
6:       select the posterior marginal estimate of the areal-unit corresponding to the original subdomain
7: Compute approximate DIC and WAIC values.
\[ \Rightarrow \text{Algorithm 3} \]

**Output:**
- **inla** object with the fitted model.
Algorithm 2 Compute mixture distributions for each log \( r_{it} \).

**Inputs:** \texttt{inla} submodel \( d \in \{1, \ldots, D\} \) containing

- \( f_d(x) \): posterior probability density function estimates of the log-risk for areal-unit \( A_{it} \).
- \( CPO_{it}^d = Pr(O_{it} = o_{it} | o_{i-1} \ldots o_{i-t}) \): conditional predictive ordinate for areal-unit \( A_{it} \).
- \( p \): number of equally spaced points at which the density is evaluated (default to 75).

**Parallel computation of mixture distributions**

1: for \( i \in \{1, \ldots, n\} \) and \( t \in \{1, \ldots, T\} \) do
2: \hspace{1em} Compute \( m(i) \): number of submodels in which the areal-unit \( A_{it} \) has been estimated (note that \( m(i) < D \))
3: \hspace{1em} Compute normalized weights
4: \hspace{2em} \( w_j = \frac{CPO_{it}^j}{\sum_j CPO_{it}^j} \), for \( j = 1, \ldots, m(i) \).
4: \hspace{1em} Compute \( f(x) = \sum_{j=1}^{m(i)} w_j f_j(x) \) evaluated at \( p \) points.

**Output:**

- Posterior marginal density estimates of log-risks.

4 Simulation study

In this section, we present two simulation studies. The first one compares the performance of our scalable model proposals with the commonly used disease mapping models described in Section 2 (denoted as Global models) in terms of risk estimation accuracy and high/low risk area detection. The second one evaluates the computational speedup offered by our modelling approach when using both parallel and/or distributed computation strategies as the number of small areas increases.

4.1 Risk estimation in high-dimensional areal data

4.1.1 Data generation

The \( n = 7907 \) municipalities of continental Spain and \( T = 25 \) time periods are used as the simulation template. For the scalable model proposals, we divide the data into \( D = 47 \) subdomains using the provinces of Spain to define a spatial partition, as this is the setting for the real data analysis presented in the next section. Under this template, a smooth risk surface is generated by sampling from a three-dimensional P-spline with 20 equally spaced knots for longitude and latitude, and 6 equally spaced knots for time. The true risk surfaces for the simulation study are shown on top of Figure 2. The simulated counts for each municipality and time point are generated from a Poisson distribution with mean \( E_{it} r_{it} \), where the number of expected cases \( E_{it} \) are fixed at value 10. A total of 50 simulated datasets have been generated.
Algorithm 3 Computations of approximate DIC and WAIC values.

Inputs:
- Posterior marginal density estimates of the risks for each $A_{it}$.
- $S$: number of samples to draw (default to 1000).

Parallel computation:
1: for $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$ do
2: Draw $S$ samples from the posterior marginal distribution of $r_{it}$.
3: Compute $\theta^s$: posterior simulations of $\mu_{it} = E_{it} r_{it}$.
4: Compute the deviance information criterion $DIC = 2 \overline{D(\theta)} - D(\bar{\theta})$, by approximating the mean deviance $\overline{D(\theta)}$ and the deviance of the mean $D(\bar{\theta})$ as

$$\overline{D(\theta)} \approx \frac{1}{S} \sum_{s=1}^{S} -2 \log(p(O|\theta^s)), \quad D(\bar{\theta}) \approx -2 \log(p(O|\bar{\theta})),$$

where $p(O|\theta)$ is the likelihood function of a Poisson distribution with mean $\theta$.
5: Approximate Watanabe-Akaike information criterion as

$$WAIC = -2 \sum_{i=1}^{n} \sum_{t=1}^{T} \log \left( \frac{1}{S} \sum_{s=1}^{S} p(O_{it}|\theta^s) \right) + 2 \sum_{i=1}^{n} \sum_{t=1}^{T} \text{Var} \left[ \log(p(O_{it}|\theta^s)) \right]$$

Output:
- Approximate DIC and WAIC values.
Figure 2: Simulation study: true risk surfaces (top) and average values of posterior median estimates of relative risks for 2nd-order neighbourhood and Type IV interaction model using the "original" merging strategy (bottom) for some selected years.
4.1.2 Results

Four different spatio-temporal models have been fitted to each simulated dataset: the Global model (Section 2) and the Disjoint, 1st-order neighbourhood and 2nd-order neighbourhood scalable model proposals (Section 3). For all these models, a BYM2 prior for the spatial random effect, a RW1 prior for the temporal random effect and the four types of interactions originally proposed by Knorr-Held (2000) for the spatio-temporal random effect have been considered.

Regarding model hyperparameters, improper uniform prior distributions are given to all the standard deviations (inverse square root of the precision parameters), and uniform prior distributions on the interval \([0, 1]\) are given to the spatial smoothing parameters of the BYM2 prior. Finally, a vague zero mean normal distribution with a precision close to zero (0.001) is given to the model intercept. All the calculations are made on a single machine with a Intel Xeon E5-2620 v4 processor and 256GB RAM (CentOS Linux release 7.3.1611 operative system), using the Gaussian approximation strategy in R-INLA (stable version INLA_22.05.07) of R-4.1.3 and simultaneously running 12 models in parallel using the \texttt{bigDM} package.

We evaluate the model performance in terms of how well relative risk are estimated by computing the mean absolute relative bias (MARB) and mean relative root mean square error (MRRMSE) for each municipality defined as

\[ \text{MARB}_i = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{50} \sum_{l=1}^{50} \left| \frac{\hat{r}_{il}^{t} - r_{it}}{r_{it}} \right|, \]

\[ \text{MRRMSE}_i = \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{1}{50} \sum_{l=1}^{50} \left( \frac{\hat{r}_{il}^{t} - r_{it}}{r_{it}} \right)^2}, \]

where \(r_{it}\) is the true generated risk, and \(\hat{r}_{il}^{t}\) is the posterior median estimate of the relative risk for areal unit \(i\) and time period \(t\) in the \(l\)-th simulation. We also compute the Interval Score (IS) for the 95% credible interval of the risks, a proper scoring rule for quantile predictions (see e.g., Gneiting and Raftery, 2007) that combines both the length and the empirical coverage of the credible interval which is defined as

\[ IS_{0.05}(r) = (u - l) + \frac{2}{0.05}(l - r)I[r < l] + \frac{2}{0.05}(r - u)I[r > u], \]

where \([l, u]\) is the 95% credible interval for the risk and \(I[\cdot]\) denotes an indicator function that penalizes the length of the credible interval if the real risk \((r)\) is not contained within that interval. For all these criteria, lower values imply better model properties.

The results of the simulation study are summarized in Table 3, where average values of Bayesian model selection criteria, risk estimation accuracy measures and computational time are displayed. For the 1st/2nd-order neighbourhood models, we compare both "mixture" and "original" merging strategies. We notice that it was computationally unfeasible to fit Type II and Type IV interaction Global models, because of the huge dimension of the spatio-temporal structure matrix \((\approx 4 \times 10^{10} \text{ elements})\) and the high number of identifiability constraints over the spatio-temporal interaction \((\approx 8000 \text{ constraints})\). In contrast, we were able to fit our scalable model proposals reducing the RAM/CPU memory usage and computational time substantially. The computational time for the Disjoint and k-order neighbourhood models are divided into running time and merging time. For the Global models only the running time is computed (we remind that these models are not scalable). The running time refers to the elapsed time for all
the submodels (which can be fitted in both parallel and/or distributed processing architectures) and the merging time corresponds to the computation of how the posterior distribution of the log-risks are combined (when necessary) and computation of approximate DIC/WAIC values in the “master node”. As expected, the complexity and computational time of the models increase as higher values of neighbourhood order are considered. On the contrary, the merging time only increases as higher neighbourhood order models are considered, as the number of areal-units for which posterior estimates must be combined increases. As it is shown in Table 3, the "original" merging strategy is less computational demanding than using mixture distributions.

Table 3: Simulation study: average values of mean deviance \( \widehat{D}(\theta) \), effective number of parameters \((p_D)\), deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC), mean absolute relative bias (MARB), mean relative root mean square error (MRRMSE), Interval Score (IS) and computational time in minutes \(T\).run: running time, \(T\).merge: merging time). For the 1st/2nd-order neighbourhood models, both "mixture" and "original" merging strategies are compared.

| Model       | Interaction | \( \widehat{D}(\theta) \) | \( p_D \) | DIC   | WAIC  | MARB  | MRRMSE | IS    | T.run | T.merge |
|-------------|-------------|-----------------------------|--------|-------|-------|-------|--------|-------|-------|---------|
| Global      | Type I      | 201406                      | 15987  | 217393| 217832| 0.0684| 0.0782 | 0.3959| 14    |         |
|             | Type II     | –                            | –      | –     | –     | –     | –      | –     | –     | –       |
|             | Type III    | 194247                      | 10683  | 204930| 204666| 0.0165| 0.0387 | 0.2816| 301   |         |
|             | Type IV     | –                            | –      | –     | –     | –     | –      | –     | –     | –       |
| Disjoint    | Type I      | 198972                      | 7563   | 206536| 206516| 0.0322| 0.0434 | 0.2582| 3     | 6       |
|             | Type II     | 200225                      | 5710   | 205934| 205965| 0.0281| 0.0419 | 0.2231| 34    | 6       |
|             | Type III    | 197112                      | 7817   | 204929| 204829| 0.0203| 0.0377 | 0.2404| 7     | 6       |
|             | Type IV     | 199103                      | 5059   | 204162| 204151| 0.0153| 0.0331 | 0.1950| 64    | 6       |

merge.strategy="mixture"

| Model       | Interaction | \( \widehat{D}(\theta) \) | \( p_D \) | DIC   | WAIC  | MARB  | MRRMSE | IS    | T.run | T.merge |
|-------------|-------------|-----------------------------|--------|-------|-------|-------|--------|-------|-------|---------|
| 1st order   | Type I      | 197900                      | 8104   | 206005| 205948| 0.0303| 0.0416 | 0.2550| 3     | 18      |
|             | Type II     | 199211                      | 6320   | 205531| 205534| 0.0261| 0.0404 | 0.2239| 53    | 18      |
|             | Type III    | 196069                      | 8366   | 204434| 204297| 0.0173| 0.0352 | 0.2400| 10    | 18      |
|             | Type IV     | 198603                      | 5243   | 203846| 203823| 0.0133| 0.0311 | 0.1974| 70    | 18      |

| Model       | Interaction | \( \widehat{D}(\theta) \) | \( p_D \) | DIC   | WAIC  | MARB  | MRRMSE | IS    | T.run | T.merge |
|-------------|-------------|-----------------------------|--------|-------|-------|-------|--------|-------|-------|---------|
| 2nd order   | Type I      | 197325                      | 8821   | 206146| 206073| 0.0312| 0.0423 | 0.2614| 4     | 32      |
|             | Type II     | 198631                      | 7070   | 205701| 205689| 0.0266| 0.0413 | 0.2341| 124   | 32      |
|             | Type III    | 195532                      | 8986   | 204518| 204353| 0.0173| 0.0356 | 0.2503| 16    | 32      |
|             | Type IV     | 198477                      | 5397   | 203874| 203848| 0.0134| 0.0312 | 0.2008| 136   | 32      |

merge.strategy="original"

| Model       | Interaction | \( \widehat{D}(\theta) \) | \( p_D \) | DIC   | WAIC  | MARB  | MRRMSE | IS    | T.run | T.merge |
|-------------|-------------|-----------------------------|--------|-------|-------|-------|--------|-------|-------|---------|
| 1st order   | Type I      | 198937                      | 7435   | 206373| 206356| 0.0322| 0.0427 | 0.2529| 3     | 8       |
|             | Type II     | 199954                      | 5836   | 205790| 205813| 0.0276| 0.0414 | 0.2215| 53    | 8       |
|             | Type III    | 196811                      | 7712   | 204523| 204424| 0.0182| 0.0357 | 0.2305| 10    | 8       |
|             | Type IV     | 198976                      | 4885   | 203861| 203846| 0.0137| 0.0312 | 0.1911| 70    | 8       |

| Model       | Interaction | \( \widehat{D}(\theta) \) | \( p_D \) | DIC   | WAIC  | MARB  | MRRMSE | IS    | T.run | T.merge |
|-------------|-------------|-----------------------------|--------|-------|-------|-------|--------|-------|-------|---------|
| 2nd order   | Type I      | 198881                      | 7592   | 206472| 206461| 0.0332| 0.0432 | 0.2524| 4     | 9       |
|             | Type II     | 199640                      | 6247   | 205887| 205905| 0.0278| 0.0421 | 0.2259| 124   | 9       |
|             | Type III    | 196473                      | 7891   | 204364| 204253| 0.0173| 0.0348 | 0.2419| 16    | 9       |
|             | Type IV     | 198897                      | 4870   | 203767| 203751| 0.0133| 0.0305 | 0.1903| 136   | 9       |
Models with a completely structured space-time interaction (Type IV interaction) perform better in this scenario in terms of model selection criteria and risk estimation accuracy measures, followed by Type III interaction models. The main reason could be that in the true risk surface considered in this scenario, the spatial pattern variability is greater than the temporal one. Specifically, the percentages of variability of the overall risk explained by each pattern is about 70% spatial, 5% temporal and 25% spatio-temporal. If the "mixture" strategy is used to combine the posterior marginal estimates of the relative risks in the border areas (that is, areal-units that are estimated in more than one submodel), slightly better results are obtained with 1st-order neighbourhood models in comparison with those considering 2nd-order neighbours. However, if the "original" strategy is used to combine the estimated risks from the different submodels, 2nd-order neighbourhood models performs better for Type III and Type IV interactions. As expected, the differences between models are more clearly stated if we compute these measures only for those areal-units located in the borders of the partition of the spatial domain (see Table A.1). As it is shown at the bottom of Figure 2, the average values of posterior median estimates of relative risks obtained with 2nd-order neighbourhood and Type IV interaction model using the "original" merging strategy are very similar to the true risk surface.

We are also interested in evaluating the models in terms of their ability to detect true high and low risk areas by calculating true positive/negative rates and false positive/negative rates. For each areal-time unit $A_{it}$, a high (low) risk area is an area where the true risk $r_{it}$ is greater (less) than one. After fitting the model, we classify an area as having high risk if the posterior probabilities that $r_{it}$ exceeds 1 is higher than a threshold value $p_0$, namely the exceedence probabilities $P(r_{it} > 1|O) > p_0$. Conversely, we classify a low risk area if the posterior probabilities that $r_{it}$ is below 1 is higher than $p_0$, i.e., $P(r_{it} < 1|O) > p_0$. Notice that these probabilities are computed from the posterior marginal distributions of the estimated relative risks. True positive rates (TPR or sensitivity) are computed as the proportion of high true risks ($r_{it} > 1$) that were correctly classified as a high risk area, while true negative rates (TNR or specificity) are computed as the proportion of low true risks ($r_{it} < 1$) that were correctly classified as a low risk area. At the same time, we are also interested in comparing the misclassification errors of the models in terms of false positive rates (FPR), i.e., the proportion of areas that are incorrectly classified as a high risk area, and false negative rates (FNR), i.e., the proportion of areas that are incorrectly classified as a low risk area.

Average values of TPR, FPR, TNR and FNR for the reference threshold values of $p_0 = 0.8, 0.9$ and 0.95 are shown in Table 4. For the 1st/2nd-order neighbourhood models, both "mixture" and "original" merging strategies are compared. We notice that that our proposed scalable models outperform the Global models in terms of high and low risk area detection. In particular, the first order neighbourhood model "original" strategy (Type IV interaction) performs the best in terms of detecting TPR. The rest of scalable models including the disjoint model (Type IV interaction) behave very similarly. In addition, if we compute these measures only for the areal-units located in the borders of the partition of the spatial domain (see Table A.2 and A.3) models with the "original" merging strategy show again better results. In general, similar values of TPR and TNR are obtained for 1st and 2nd order neighbourhood models using both merging strategies. In terms of false positive and false negative rates, although slightly better results are obtained when the "mixture" strategy is used, very low values are obtained in all cases.
Table 4: Simulation study: average values of true/false positive rates and true/false negative rates for the reference threshold values of $p_0 = 0.8, 0.9$ and $0.95$, based on posterior exceedence probabilities $P(r_{it} > 1|O)$ and $P(r_{it} < 1|O)$, respectively. For the 1st/2nd-order neighbourhood models, both "mixture" and "original" merging strategies are compared.

| Model   | Space-time interaction | True Positive Rate | True Negative Rate | False Positive Rate | False Negative Rate |
|---------|------------------------|--------------------|-------------------|---------------------|---------------------|
|         | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ |
| Global  | Type I     | 0.5932 | 0.4349 | 0.3212 | 0.6902 | 0.5453 | 0.4219 | 0.0142 | 0.0024 | 0.0004 | 0.0282 | 0.0130 | 0.0069 |
|         | Type II    | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
|         | Type III   | 0.7414 | 0.6211 | 0.5226 | 0.8122 | 0.6937 | 0.5892 | 0.0025 | 0.0003 | 0.0000 | 0.0044 | 0.0006 | 0.0001 |
|         | Type IV    | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Disjoint| Type I     | 0.7742 | 0.6706 | 0.5836 | 0.8368 | 0.7520 | 0.6660 | 0.0119 | 0.0043 | 0.0018 | 0.0181 | 0.0074 | 0.0032 |
|         | Type II    | 0.8021 | 0.7160 | 0.6411 | 0.8557 | 0.7864 | 0.7209 | 0.0111 | 0.0038 | 0.0014 | 0.0182 | 0.0078 | 0.0036 |
|         | Type III   | 0.7764 | 0.6721 | 0.5868 | 0.8403 | 0.7509 | 0.6705 | 0.0045 | 0.0009 | 0.0002 | 0.0080 | 0.0019 | 0.0005 |
|         | Type IV    | 0.8256 | 0.7449 | 0.6754 | 0.8747 | 0.8065 | 0.7408 | 0.0048 | 0.0011 | 0.0003 | 0.0079 | 0.0020 | 0.0006 |
| merge.strategy="mixture" | | | | | | | | | | | | |
| 1st order| Type I     | 0.7685 | 0.6601 | 0.5088 | 0.8345 | 0.7421 | 0.6531 | 0.0088 | 0.0026 | 0.0009 | 0.0148 | 0.0054 | 0.0021 |
|         | Type II    | 0.7941 | 0.7028 | 0.6235 | 0.8531 | 0.7751 | 0.7039 | 0.0082 | 0.0024 | 0.0007 | 0.0147 | 0.0055 | 0.0023 |
|         | Type III   | 0.7684 | 0.6602 | 0.5710 | 0.8397 | 0.7413 | 0.6529 | 0.0027 | 0.0004 | 0.0001 | 0.0051 | 0.0009 | 0.0002 |
|         | Type IV    | 0.8208 | 0.7382 | 0.6681 | 0.8762 | 0.8040 | 0.7349 | 0.0033 | 0.0006 | 0.0001 | 0.0056 | 0.0012 | 0.0003 |
| 2nd order| Type I     | 0.7605 | 0.6451 | 0.5480 | 0.8296 | 0.7272 | 0.6327 | 0.0074 | 0.0017 | 0.0004 | 0.0138 | 0.0046 | 0.0019 |
|         | Type II    | 0.7833 | 0.6860 | 0.6004 | 0.8480 | 0.7611 | 0.6816 | 0.0069 | 0.0018 | 0.0005 | 0.0142 | 0.0050 | 0.0020 |
|         | Type III   | 0.7614 | 0.6503 | 0.5573 | 0.8363 | 0.7305 | 0.6347 | 0.0024 | 0.0003 | 0.0001 | 0.0050 | 0.0008 | 0.0002 |
|         | Type IV    | 0.8172 | 0.7336 | 0.6633 | 0.8756 | 0.8012 | 0.7294 | 0.0031 | 0.0006 | 0.0001 | 0.0057 | 0.0012 | 0.0002 |
| merge.strategy="original" | | | | | | | | | | | | |
| 1st order| Type I     | 0.7736 | 0.6698 | 0.5832 | 0.8406 | 0.7557 | 0.6099 | 0.0111 | 0.0037 | 0.0014 | 0.0179 | 0.0071 | 0.0029 |
|         | Type II    | 0.7992 | 0.7113 | 0.6354 | 0.8573 | 0.7859 | 0.7186 | 0.0099 | 0.0032 | 0.0010 | 0.0174 | 0.0072 | 0.0032 |
|         | Type III   | 0.7749 | 0.6707 | 0.5863 | 0.8447 | 0.7352 | 0.6999 | 0.0034 | 0.0006 | 0.0001 | 0.0063 | 0.0013 | 0.0003 |
|         | Type IV    | 0.8260 | 0.7461 | 0.6782 | 0.8798 | 0.8117 | 0.7402 | 0.0037 | 0.0008 | 0.0002 | 0.0064 | 0.0014 | 0.0003 |
| 2nd order| Type I     | 0.7688 | 0.6623 | 0.5753 | 0.8405 | 0.7533 | 0.6662 | 0.0109 | 0.0033 | 0.0011 | 0.0181 | 0.0070 | 0.0029 |
|         | Type II    | 0.7910 | 0.6994 | 0.6210 | 0.8546 | 0.7793 | 0.7080 | 0.0092 | 0.0027 | 0.0008 | 0.0172 | 0.0070 | 0.0031 |
|         | Type III   | 0.7717 | 0.6671 | 0.5824 | 0.8449 | 0.7301 | 0.6636 | 0.0027 | 0.0004 | 0.0001 | 0.0054 | 0.0010 | 0.0002 |
|         | Type IV    | 0.8245 | 0.7445 | 0.6775 | 0.8812 | 0.8124 | 0.7462 | 0.0033 | 0.0006 | 0.0001 | 0.0059 | 0.0012 | 0.0003 |
4.2 Numerical simulation

In this section we want to evaluate the computational gain offered by the scalable modelling approach against the Global model as the number of small areas increases. Specifically, we simulate a regular grid map with number of areas equal to $n = 256, 1024$ and $4096$, while the number of time points have been fixed to $T = 25$ to imitate the real data analysis. For each template, spatially structured (CAR), temporally structured (RW1) and completely structured spatio-temporal (Type IV) random effects are generated from the corresponding structure matrices to define a log-risk surface (see Equation (1)). Finally, we simulate counts for each areal-unit from a Poisson distribution as described in the previous section. To fit the scalable models, a $4 \times 4$ regular grid is used to define the partition of the spatial domain, so that a total of $D = 16$ local spatio-temporal models are fitted. These models are distributed over 4 machines with 4 models running in parallel for each machine using the bigDM package.

In Figure 3, we show the total runtime of the different models when varying the number of spatial areas. As we increase the dimension of the spatial domain, the Global model quickly becomes computationally prohibitive. For $n = 256$ areas the total running time is about 80 minutes, while for $n = 1024$ areas the computational time exceeds 120 hours. For larger area sizes considered here, computation fails due to very high RAM memory usage. On the other hand, we can fit the Disjoint and k-order neighbourhood models for $n = 256$ areas in total running times between 2-6 minutes, for $n = 1024$ areas in times between 6-36 minutes, and for $n = 4096$ areas in running times between 5-16 hours.

![Figure 3: Numerical simulation: computational time (in log10 scale) vs number of small areas for the Global model and our scalable modelling proposals.](image-url)
5 Data analysis: lung cancer mortality in Spain

We illustrate and compare all the approaches described in this paper by modelling the spatio-temporal evolution of male lung cancer mortality data in the \( n = 7907 \) municipalities of continental Spain (excluding Balearic and Canary Islands and the autonomous cities of Ceuta and Melilla) during the period 1991-2015. According to recent studies (Ferlay et al., 2018), lung cancer was the leading cause of cancer deaths among the male population and the second cause among the female population in Europe in 2018, representing 24.8% and 14.2% of all cancer deaths, respectively. It also was the leading cause of cancer related deaths in Spain for both sexes in 2017, representing the 19.5% of cancer mortality (Spanish Society of Medical Oncology (Sociedad Española de Oncología Médica), 2020). One of the main causes is that Spain is a country with a traditionally high tobacco consumption, with a smoking rate of over 32% of the population at the beginning of the century (Remon et al., 2021).

A total of 378,720 lung cancer deaths (corresponding to International Classification of Diseases-10 codes C33-C34) were registered for male population in the municipalities of continental Spain during the period 1991-2015 (that account for around 26% of all malignant tumours for the target population during our study period), where the number of observed deaths per areal-time unit varies from 0 to 1247 (with a mean value of 1.9). The number of expected cases were computed using the indirect (internal) standardization method with 5-years age groups as standardization variable. The number of expected deaths per areal-time unit varies from 0 to 1332 (with a mean value of 1.9). A brief summary of the number of observed deaths, expected deaths and standardized mortality ratios for the provincial capital municipalities during the whole period is displayed in Table 5.

For the Disjoint and 1st/2nd-order neighbourhood models presented in this section, we distribute the models over 7 machines with Intel Xeon E5-2620 v4 processors and 256GB RAM on each machine (CentOS Linux release 7.3.1611 operative system), using the simplified Laplace

| Municipality | Obs. | Exp. | SMR | Municipality | Obs. | Exp. | SMR | Municipality | Obs. | Exp. | SMR |
|--------------|------|------|-----|-------------|------|------|-----|-------------|------|------|-----|
| Ávila         | 381  | 467.2| 0.815| Salamanca    | 1568 | 1621.7| 0.967| Alicante     | 2899 | 2634.2| 1.101|
| Segovia       | 443  | 533.4| 0.831| Palencia     | 760  | 785.0 | 0.968| Barcelona    | 18161| 16434.4| 1.105|
| Burgos        | 1335 | 1572.0| 0.849| Girona       | 644  | 663.5 | 0.971| Almería      | 1491 | 1324.4| 1.126|
| Vitoria       | 1795 | 2053.5| 0.874| Lugo         | 886  | 903.0 | 0.981| A Coruña     | 2603 | 2385.2| 1.129|
| Logroño       | 1098 | 1231.8| 0.891| San Sebastián| 1750 | 1781.0| 0.983| Zaragoza     | 6875 | 6083.2| 1.130|
| Guadalajara  | 576  | 644.9 | 0.893| Madrid       | 28750| 29048.5| 0.990| Santander    | 2049 | 1786.6| 1.147|
| Cuenca        | 405  | 452.6 | 0.895| Valladolid   | 3034 | 3060.3| 0.991| Sevilla      | 6605 | 5667.2| 1.165|
| Jaén          | 784  | 876.3 | 0.895| Tarragona    | 1063 | 1066.8| 0.996| Valencia     | 8261 | 7046.9| 1.172|
| Soria         | 322  | 375.2 | 0.901| Pamplona     | 1837 | 1804.1| 1.018| Ciudad Real  | 610  | 513.7 | 1.187|
| Albacete      | 1103 | 1208.5| 0.913| Murcia       | 3048 | 2982.6| 1.022| Oviedo       | 2450 | 2026.9| 1.209|
| Ourense       | 1020 | 1104.3| 0.924| Pontevedra   | 669  | 646.7 | 1.034| Málaga       | 5122 | 4213.2| 1.216|
| Zamora        | 618  | 668.4 | 0.925| Toledo       | 613  | 579.6 | 1.058| Cáceres      | 839  | 653.6 | 1.284|
| Granada       | 1940 | 2094.7| 0.926| Lérida       | 1164 | 1094.9| 1.063| Huelva       | 1447 | 1089.6| 1.328|
| Teruel        | 298  | 318.0 | 0.937| Cordoba      | 2767 | 2591.8| 1.068| Badajoz      | 1445 | 1055.5| 1.369|
| Huesca        | 449  | 472.7 | 0.950| Castellón    | 1427 | 1320.7| 1.080| Cádiz        | 1637 | 1164.4| 1.406|
| León          | 1406 | 1455.5| 0.966| Bilbao       | 4053 | 3702.6| 1.095|             |      |       |     |
Table 6: Model selection criteria and computational time (in minutes) for models fitted using the simplified Laplace approximation strategy of INLA and the "original" merging strategy.

| Model          | Interaction | $\bar{D}(\theta)$ | $p_D$  | DIC    | WAIC   | T.run | T.merge | T.total |
|----------------|-------------|--------------------|--------|--------|--------|-------|---------|---------|
| Global         | Type I      | 144680             | 2984   | 147664 | 147696 | 663   | -       | 663     |
|                | Type II     | -                  | -      | -      | -      | -     | -       | -       |
|                | Type III    | 144467             | 2968   | 147435 | 147458 | 3846  | -       | 3846    |
|                | Type IV     | -                  | -      | -      | -      | -     | -       | -       |
| Disjoint       | Type I      | 143154             | 3999   | 147154 | 147161 | 10    | 6       | 16      |
|                | Type II     | 143175             | 3801   | 146976 | 147045 | 218   | 6       | 224     |
|                | Type III    | 143101             | 4015   | 147116 | 147161 | 22    | 6       | 27      |
|                | Type IV     | 143131             | 3753   | 146884 | 146965 | 259   | 6       | 264     |
| 1st order      | Type I      | 143269             | 3824   | 147094 | 147112 | 14    | 8       | 22      |
| neighbourhood  | Type II     | 143255             | 3671   | 146926 | 146997 | 548   | 8       | 557     |
|                | Type III    | 143497             | 3562   | 147058 | 147123 | 34    | 8       | 42      |
|                | Type IV     | 143370             | 3458   | 146828 | 146910 | 636   | 8       | 644     |
| 2nd order      | Type I      | 143500             | 3603   | 147103 | 147138 | 19    | 10      | 28      |
| neighbourhood  | Type II     | 143366             | 3566   | 146932 | 147009 | 1740  | 10      | 1750    |
|                | Type III    | 143731             | 3331   | 147062 | 147130 | 59    | 10      | 68      |
|                | Type IV     | 143523             | 3307   | 146830 | 146912 | 1879  | 10      | 1889    |

$\bar{D}(\theta)$: mean deviance, $p_D$: effective number of parameters
DIC: deviance information criterion, WAIC: Watanabe-Akaike information criterion
T.run: running time, T.merge: merging time, T.total: running + merging time

approximation strategy in R-INLA (stable version INLA 22.05.07) of R-4.1.3 and simultaneously running 5 models in parallel on each machine using the bigDM package. Again, it was not possible to fit Type II and Type IV interaction Global models using a single machine with the described characteristics. Results in terms of model selection criteria and computational time are shown in Table 6. For the scalable models only the results regarding the "original" merging strategy are shown, since the simulation study shows that this procedure outperforms the "mixture" strategy in terms of risk estimation accuracy and high/low risk area detection.

It can be seen that both DIC and WAIC model selection criteria support Type IV and Type II interaction effects, which precisely are those that cannot be fitted with the Global model. Besides the computational gain, the scalable model proposals are better supported by fit measures. In particular, 1st-order neighbourhood models shows slightly better performance. Maps with the posterior median estimates of relative risks and posterior exceedence probabilities $P(r_{it} > 1|O)$ obtained with the 1st-order neighbourhood model considering a Type IV interaction for the spatio-temporal random effect are plotted at Figure 4. The estimated risk surfaces are consistent with those described by López-Abente et al. (2014), where the geographical pattern of lung cancer mortality data in Spain at municipality level was analysed using spatial models.
Figure 4: Maps of posterior median estimates of relative risks $r_{it}$ (top) and posterior exceedence probabilities $P(r_{it} > 1|O)$ (bottom) for the 1st-order neighbourhood model considering a BYM2 conditional autoregressive prior for space, RW1 prior for time and Type IV interaction for the spatio-temporal effect.
6 Conclusions

The use of spatial and spatio-temporal hierarchical models for regional data are crucial in areas such as cancer epidemiology, since they allow to obtain reliably incidence or mortality risk estimates of cancer in small areas, avoiding the huge variability of classical risk estimation measures such as the standardized mortality ratios or the crude rates. Research in this area has been very fruitful in recent decades and numerous statistical models have been proposed to study the geographic distribution of cancer and its evolution in time, as well as the underlying spatio-temporal patterns. However, the scalability of these models, i.e., their use when the number of space-time domains increases significantly, has not been studied in depth yet. For that reason, the pragmatic, simple, and useful methodology proposed in this paper aims to provide alternative modelling approaches to disease mapping models commonly used when analysing high-dimensional spatio-temporal data.

Despite the enormous expansion of modern computers and the development of new software and estimation techniques to make fully Bayesian inference, dealing with massive data is still computationally challenging. Our proposal is based on the idea of “divide-and-conquer” so that local spatio-temporal models can be simultaneously fitted. Adapting this idea to the context of disease mapping seems to be very appropriate when the number of small areas is very large for three main reasons: (1) it is a natural and “simple” strategy, (2) the larger the spatial domain is, the less likely it is that the data are stationary across the whole map, and (3) it provides a scalable modelling scheme that substantially reduces the RAM/CPU memory usage and computational time.

Our simulation study indicates that the proposed methodology provides reliable risk estimates with a substantial reduction in computational time. Furthermore, we observe that our model proposals perform better in detecting high/low risk areas, by obtaining higher true positive and true negative rates than when considering the usual spatio-temporal CAR models, avoiding false alarms. Regarding the merging strategy of the areas belonging to different subdomains, we compare the use of mixture distributions to combine the posterior marginal density functions against using the posterior marginal estimate of the areal-unit corresponding to the original subdomain. Our simulation study shows that the latter strategy (denoted as “original” merging strategy) reduces computational time while providing better results in terms of risk estimation accuracy and true positive/negative rates. On the other hand, in some cases it may not be sufficient to use first-order neighbours to avoid the boundary effect caused by the division of the whole study region into smaller subdomains. We have additionally analyzed the advantages of our scalable model proposal in terms of computational complexity as the number of small areas increases. Our numerical simulation study shows a substantial reduction in computational time in comparison with the Global models. Finally, lung cancer mortality data in the municipalities of Spain during the period 1991-2015 have been analyzed to illustrate the new model proposals, using the administrative division of continental Spain in 47 provinces to define the partition of the spatial domain. Doing so, we are able to fit a CAR model that accounts for both spatial and temporal dependence by including completely structured space-time interaction random effects (commonly denoted as Type IV interaction), which was computationally unfeasible to fit when considering non-scalable models.

The methods and algorithms proposed in this work are implemented in the open-source R package bigDM (https://cran.r-project.org/web/packages/bigDM/index.html). This package allows the user to adapt the modelling scheme to their own processing architecture by
performing both parallel and/or distributed computation strategies to speed up computations by using the *future* package. Model fitting and inference is carried out using INLA methodology through the *R-INLA*, as it is now a well-known Bayesian approximation technique, computationally efficient and easy for practitioners to handle. Very recently, promising research in a hybrid approximate method that uses the Laplace method with a low-rank Variational Bayes correction to the posterior mean has been released (van Niekerk and Rue, 2021; van Niekerk et al., 2022). This new approximation technique has been shown to provide accurate results with a superior computational efficiency and scalability than the classic integrated nested Laplace approximations. Currently, it is implemented in *R-INLA* as an experimental mode, but as stated by the developers of this technique, it will presumably be enabled by default in the near future. When the moment comes, we plan to adapt our *bigDM* package to be compatible with this new avenue for Bayesian inference with INLA.

Finally, we are currently working on extending our Bayesian modelling proposal to ecological regression models that takes into account the spatial and/or spatio-temporal confounding issues between fixed and random effects (Adin et al., 2021), as well as to high-dimensional multivariate disease mapping models in which several diseases are jointly analysed (Vicente et al., 2021).

**Acknowledgements**

This research has been supported by the project PID2020-113125RB-I00/MCIN/AEI/10.13039/501100011033. It has also been partially funded by the Public University of Navarra (project PJUPNA20001). We would like to thank Andrea Riebler and James Hodges for their useful comments that have contributed to the improvement of this paper.

**References**

Adin, A., Goicoa, T., Hodges, J. S., Schnell, P. M., and Ugarte, M. D. (2021). Alleviating confounding in spatio-temporal areal models with an application on crimes against women in India. *Statistical Modelling (published online on May 31, 2021).*

Adin, A., Orozco-Acosta, E., and Ugarte, M. D. (2022). *bigDM: Scalable Bayesian Disease Mapping Models for High-Dimensional Data.* R package version 0.4.2.

Appel, M. and Pebesma, E. (2020). Spatiotemporal multi-resolution approximations for analyzing global environmental data. *Spatial Statistics,* 38:100465.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics,* 10(6):e1443.

Banerjee, S. (2017). High-dimensional Bayesian geostatistics. *Bayesian Analysis,* 12(2):583.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data,* volume 135 of *Monographs on Statistics and Applied Probability.* CRC Press, Boca Raton, FL, second edition.

Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in R using futures. *The R Journal (accepted article).*
Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1):1–32.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association, 111*(514):800–812.

Datta, A., Banerjee, S., Hodges, J. S., and Gao, L. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models. *Bayesian Analysis, 14*(4):1221.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics, 26*(2):403–413.

Dean, C. B., Ugarte, M. D., and Militino, A. F. (2001). Detecting interaction between random regions and fixed age effects in disease mapping. *Biometrics, 57*(1):197–202.

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., and Bray, F. (2018). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer, 103*:356–387.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., and Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics, 28*(2):401–414.

Gao, L., Datta, A., and Banerjee, S. (2021). Hierarchical Multivariate Directed Acyclic Graph Auto-Regressive (MDAGAR) models for spatial diseases mapping. *arXiv preprint arXiv:2102.02911*.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association, 102*(477):359–378.

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.

Guan, Y. and Haran, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics, 27*(4):701–714.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics, 24*(3):398–425.
Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159.

Jurek, M. and Katzfuss, M. (2022). Hierarchical sparse cholesky decomposition with applications to high-dimensional spatio-temporal filtering. *Statistics and Computing (published online on January 18, 2022)*, 32(15).

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112(517):201–214.

Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, 36(1):124–141.

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., Banerjee, S., Haining, R. P., and Ugarte, M. D. (2016). *Handbook of Spatial Epidemiology*. CRC Press.

Leroux, B. G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Halloran, M. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer-Verlag: New York.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When Gaussian process meets big data: A review of scalable GPs. *IEEE transactions on neural networks and learning systems*, 31(11):4405–4423.

López-Abente, G., Aragonés, N., Pérez-Gómez, B., Pollán, M., García-Pérez, J., Ramis, R., and Fernández-Navarro, P. (2014). Time trends in municipal distribution patterns of cancer mortality in spain. *BMC Cancer*, 14(1):1–15.

Orozco-Acosta, E., Adin, A., and Ugarte, M. D. (2021). Scalable Bayesian modelling for smoothing disease risks in large spatial data sets using INLA. *Spatial Statistics*, 41:100496.

Peruzzi, M., Banerjee, S., and Finley, A. O. (2020). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *Journal of the American Statistical Association*, pages 1–14.

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):175–184.
Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd international workshop on distributed statistical computing, volume 124, pages 1–10. Vienna, Austria.

Remon, J. et al. (2021). Lung cancer in Spain. Journal of Thoracic Oncology, 16(2):197–204.

Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Statistical Methods in Medical Research, 25(4):1145–1165.

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.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. Annual Review of Statistics and Its Application, 4:395–421.

Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. Journal of the Royal Statistical Society: Series C (Applied Statistics), 60(2):261–279.

Shen, W. and Louis, T. A. (2000). Triple-goal estimates for disease mapping. Statistics in Medicine, 19(17-18):2295–2308.

Sørbye, S. H. and Rue, H. (2014). Scaling intrinsic Gaussian Markov field priors in spatial modelling. Spatial Statistics, 8:39–51.

Spanish Society of Medical Oncology (Sociedad Española de Oncología Médica) (2020). Las cifras del cáncer en España 2020.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS user manual.

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.

Sun, Y., Li, B., and Genton, M. G. (2012). Geostatistics for Large Datasets. In Advances and challenges in space-time modelling of natural events, pages 55–77. Springer.

Urdangarin, A., Goicoc, T., and Ugarte, M. D. (2022). Space-time interactions in Bayesian disease mapping with recent tools: Making things easier for practitioners. Statistical Methods in Medical Research, 31(6):1085–1103.

van Niekerk, J., Krainski, E., Rustand, D., and Rue, H. (2022). A new avenue for Bayesian inference with INLA. arXiv preprint.

van Niekerk, J. and Rue, H. (2021). Correcting the Laplace Method with Variational Bayes. arXiv preprint.
Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312.

Vicente, G., Goicoa, T., and Ugarte, M. D. (2021). Multivariate Bayesian spatio-temporal P-spline models to analyze crimes against women. *Biostatistics (published online on Dec 27, 2021)*.

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.

Zammit-Mangion, A. and Rougier, J. (2020). Multi-scale process modelling and distributed computation for spatial data. *Statistics and Computing*, 30(6):1609–1627.
## Appendix

Table A.1: Simulation study: average values of mean absolute relative bias (MARB), mean relative root mean square error (MRRMSE) and Interval Score (IS) for the 1st/2nd-order neighbourhood models **computed only for the border areas.**

| Model                  | Interaction | merge.strategy="mixture" |         |         | merge.strategy="original" |         |         |
|------------------------|-------------|---------------------------|---------|---------|---------------------------|---------|---------|
|                        |             | MARB  | MRRMSE | IS      | MARB  | MRRMSE | IS      |
| 1st order              | Type I      | 0.0304 | 0.0430 | 0.2733  | 0.0375 | 0.0471 | 0.2656  |
|                        | Type II     | 0.0264 | 0.0410 | 0.2394  | 0.0318 | 0.0446 | 0.2305  |
|                        | Type III    | 0.0178 | 0.0369 | 0.2759  | 0.0211 | 0.0389 | 0.2405  |
|                        | Type IV     | 0.0141 | 0.0322 | 0.2140  | 0.0154 | 0.0326 | 0.1906  |
| 2nd order              | Type I      | 0.0323 | 0.0438 | 0.2734  | 0.0359 | 0.0454 | 0.2576  |
|                        | Type II     | 0.0276 | 0.0422 | 0.2432  | 0.0298 | 0.0435 | 0.2287  |
|                        | Type III    | 0.0181 | 0.0369 | 0.2715  | 0.0181 | 0.0356 | 0.2409  |
|                        | Type IV     | 0.0141 | 0.0320 | 0.2083  | 0.0138 | 0.0308 | 0.1900  |
Table A.2: Simulation study: average values of true and false positive rates for the reference threshold values of $p_0 = 0.8, 0.9$ and 0.95, based on posterior exceedence probabilities $P(r_{it} > 1|O)$. Results for the 1st/2nd-order neighbourhood models **computed only for the border areas.**

|                  | merge.strategy=“mixture” | merge.strategy=“original” |
|------------------|--------------------------|---------------------------|
|                  | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ |
| **True Positive Rate** |                      |                           |
| 1st order        |                      |                           |
| neighbourhood    |                      |                           |
| Type I           | 0.7598               | 0.6490                   | 0.5494       | 0.7789       | 0.6848       | 0.6024       |
| Type II          | 0.7864               | 0.6939                   | 0.6104       | 0.8051       | 0.7255       | 0.6549       |
| Type III         | 0.7609               | 0.6513                   | 0.5545       | 0.7849       | 0.6901       | 0.6106       |
| Type IV          | 0.8122               | 0.7302                   | 0.6599       | 0.8313       | 0.7591       | 0.6972       |
| 2nd order        |                      |                           |
| neighbourhood    |                      |                           |
| Type I           | 0.7601               | 0.6459                   | 0.5442       | 0.7745       | 0.6757       | 0.5916       |
| Type II          | 0.7840               | 0.6889                   | 0.6023       | 0.7973       | 0.7121       | 0.6379       |
| Type III         | 0.7657               | 0.6583                   | 0.5636       | 0.7836       | 0.6875       | 0.6071       |
| Type IV          | 0.8185               | 0.7384                   | 0.6703       | 0.8312       | 0.7572       | 0.6950       |
| **False Positive Rate** |                |                           |
| 1st order        |                      |                           |
| neighbourhood    |                      |                           |
| Type I           | 0.0054               | 0.0011                   | 0.0002       | 0.0142       | 0.0052       | 0.0021       |
| Type II          | 0.0053               | 0.0012                   | 0.0003       | 0.0119       | 0.0043       | 0.0015       |
| Type III         | 0.0021               | 0.0003                   | 0.0000       | 0.0045       | 0.0009       | 0.0002       |
| Type IV          | 0.0027               | 0.0004                   | 0.0001       | 0.0044       | 0.0009       | 0.0002       |
| 2nd order        |                      |                           |
| neighbourhood    |                      |                           |
| Type I           | 0.0062               | 0.0013                   | 0.0003       | 0.0125       | 0.0042       | 0.0016       |
| Type II          | 0.0060               | 0.0015                   | 0.0004       | 0.0100       | 0.0031       | 0.0010       |
| Type III         | 0.0024               | 0.0004                   | 0.0001       | 0.0029       | 0.0005       | 0.0001       |
| Type IV          | 0.0029               | 0.0005                   | 0.0001       | 0.0031       | 0.0006       | 0.0001       |
Table A.3: Simulation study: average values of true and false negative rates for the reference threshold values of $p_0 = 0.8, 0.9$ and $0.95$, based on posterior exceedence probabilities $P(r_{it} < 1|O)$. Results for the 1st/2nd-order neighbourhood models computed only for the border areas.

|                | merge.strategy="mixture" | merge.strategy="original" |
|----------------|---------------------------|----------------------------|
|                | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ |
| 1st order      |                |                |                |                |                |                |
| neighbourhood  |                |                |                |                |                |                |
| Type I         | 0.8063        | 0.6923         | 0.5954         | 0.8291        | 0.7431         | 0.6580         |
| Type II        | 0.8337        | 0.7359         | 0.6523         | 0.8499        | 0.7768         | 0.7084         |
| Type III       | 0.8182        | 0.6983         | 0.5943         | 0.8367        | 0.7427         | 0.6576         |
| Type IV        | 0.8618        | 0.7774         | 0.6984         | 0.8753        | 0.8060         | 0.7405         |
| 2nd order      |                |                |                |                |                |                |
| neighbourhood  |                |                |                |                |                |                |
| Type I         | 0.8139        | 0.7008         | 0.6030         | 0.8331        | 0.7470         | 0.6622         |
| Type II        | 0.8388        | 0.7424         | 0.6571         | 0.8507        | 0.7751         | 0.7043         |
| Type III       | 0.8293        | 0.7148         | 0.6109         | 0.8445        | 0.7494         | 0.6622         |
| Type IV        | 0.8725        | 0.7938         | 0.7178         | 0.8824        | 0.8137         | 0.7476         |

|                | merge.strategy="mixture" | merge.strategy="original" |
|----------------|---------------------------|----------------------------|
|                | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ | $p_0 = 0.8$ | $p_0 = 0.9$ | $p_0 = 0.95$ |
| 1st order      |                |                |                |                |                |                |
| neighbourhood  |                |                |                |                |                |                |
| Type I         | 0.0111        | 0.0029         | 0.0007         | 0.0225        | 0.0091         | 0.0036         |
| Type II        | 0.0108        | 0.0029         | 0.0007         | 0.0209        | 0.0091         | 0.0042         |
| Type III       | 0.0040        | 0.0005         | 0.0001         | 0.0084        | 0.0019         | 0.0004         |
| Type IV        | 0.0043        | 0.0007         | 0.0001         | 0.0073        | 0.0016         | 0.0004         |
| 2nd order      |                |                |                |                |                |                |
| neighbourhood  |                |                |                |                |                |                |
| Type I         | 0.0124        | 0.0030         | 0.0007         | 0.0199        | 0.0071         | 0.0025         |
| Type II        | 0.0123        | 0.0032         | 0.0009         | 0.0176        | 0.0066         | 0.0027         |
| Type III       | 0.0047        | 0.0007         | 0.0001         | 0.0055        | 0.0010         | 0.0002         |
| Type IV        | 0.0052        | 0.0009         | 0.0002         | 0.0055        | 0.0011         | 0.0002         |