Computer Methods and Programs in Biomedicine 231 (2023) 107403

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Big problems in spatio-temporal disease mapping: Methods and software

Erick Orozco-Acosta[a,b], Aritz Adin[a,b], María Dolores Ugarte[a,b,*]

[a] Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Campus de Arrosadia, 31006 Pamplona, Spain
[b] Institute for Advanced Materials and Mathematics (InaMat2), Public University of Navarre, Campus de Arrosadia, 31006 Pamplona, Spain

ARTICLE INFO

Article history:
Received 11 October 2022
Revised 12 January 2023
Accepted 1 February 2023

Keywords:
Cancer epidemiology
Laplace approximations
Massive data
Non-stationary models
Scalable modelling

ABSTRACT

Background and objective: Fitting spatio-temporal models for areal data is crucial in many fields such as cancer epidemiology. However, when data sets are very large, many issues arise. The main objective of this paper is to propose a general procedure to analyze high-dimensional spatio-temporal areal data, with special emphasis on mortality/incidence relative risk estimation.

Methods: We present a pragmatic and simple idea that permits hierarchical spatio-temporal models to be fitted 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 or even unfeasible.

Results: Using simulated and real data, we show that our method outperforms classical global models. We implement the methods and algorithms that we develop in the open-source R package `bigDM` where specific vignettes have been included to facilitate the use of the methodology for non-expert users.

Conclusions: Our scalable methodology proposal provides reliable risk estimates when fitting Bayesian hierarchical spatio-temporal models for high-dimensional data.

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 analyze 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 [1,2] 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. [3] and Banerjee [4] 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. [5] and Liu et al. [6] for recent reviews and comparisons). Some other recent methods to deal with massive datasets are described below. Appel and Pebesma [7] provide an extension to the multi-resolution approximation approach [8] 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 [9] propose an approximate inference scalable algorithm for multi-scale process modelling by using the stochastic partial differential equation approach [10]. Both

* Corresponding author.
E-mail addresses: erick.orozco@unavarra.es (E. Orozco-Acosta), aritz.adin@unavarra.es (A. Adin), lola@unavarra.es (M.D. Ugarte).

https://doi.org/10.1016/j.cmpb.2023.107403

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[11]. This approximation can be regarded as a special case of the Gaussian Markov random field approximations [12] 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. [13] propose alternative formulations of Bayesian nearest neighbour Gaussian process models developed by Datta et al. [14] to substantially improve computational efficiency; Peruzzi et al. [15] develop a meshed Gaussian process with a novel partitioning and graph design based on domain tessellations while [16] propose a novel sparse general Vecchia approximation algorithm which ensures computational feasibility for large spatial datasets; Jurek and Katzfuss [17] 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 [18]. As outlined by Shen and Louis [19], 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 [20] 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 [21]. This model is similar to the one proposed by Hughes and Haran [22], where the decomposition is performed based on the Moran operator. Datta et al. [23] 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. [24]. Very recently, a scalable Bayesian spatial model has been proposed by Orozco-Acosta et al. [25] 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 a high-dimensional spatio-temporal disease mapping context by extending the methodology described in Orozco-Acosta et al. [25]. 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 non-stationary models to be defined, 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. Section 4 describes the implementation of our proposed methodology in the R package `bigDM`. In Section 5, we conduct a simulation study based on a template of almost 8000 municipalities of continental Spain and 25 time periods to compare the new scalable methods with previous proposals. In addition, we provide a numerical simulation to evaluate the computational gain offered by our modelling approach when the number of small areas increases. In Section 6 we use the new model proposal to analyze lung cancer mortality data in Spanish municipalities. The paper ends with a discussion.

2. Background: 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

$$E_{it} = \sum_{j=1}^{J} N_{it} \frac{O_j}{N_j} \quad \text{for } i = 1, \ldots, n;\ t = 1, \ldots, T,$$

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 with high-dimensional data. This makes 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 on 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,

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

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

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 [26] are probably 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

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

where α is an intercept representing the overall log-risk, ξ_i is a spatial random effect with CAR prior distribution, γ_t is a temporally structured random effect that follows a random walk prior distribution, and δ_{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 structured matrices.

A modification of the Dean et al. [27] model proposed by Riebler et al. [28], hereafter called BYM2 model, has been considered as the prior distribution for the spatial random effects $\xi = (\xi_1, \ldots, \xi_n)'$, so that

$$\xi = \frac{1}{\sqrt{\tau_\xi}}\left(\sqrt{1-\lambda_\xi}\mathbf{v} + \sqrt{\lambda_\xi}\mathbf{u}_\star\right),$$

where τ_ξ is a precision parameter, $\lambda_\xi \in [0,1]$ is a spatial smoothing parameter, $\mathbf{v}$ is the vector of unstructured random effects and $\mathbf{u}_\star$ is the scaled intrinsic CAR model with generalized variance equal to one. Note that the variance of ξ is expressed as a weighted average of the covariance matrices of the unstructured and structured spatial components (unlike the CAR model proposed by Leroux et al. [29] which considers a weighted combination of the precision matrices), i.e.,

$$\xi \sim N(\mathbf{0}, \mathbf{Q}_\xi^\star), \quad \text{with } \mathbf{Q}_\xi^\star = \tau_\xi^{-1}[(1-\lambda_\xi)\mathbf{I}_n + \lambda_\xi \mathbf{R}_\star^-],$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathbf{R}_\star^-$ indicates the generalised inverse of the scaled spatial structure matrix corresponding to the undirected graph of the regions under study (see, e.g., Sørbye and Rue [30]). Recall that the spatial structure matrix is defined as $\mathbf{R}_\xi = \mathbf{D}_W - \mathbf{W}$, where $\mathbf{D}_W = diag(w_{1+}, \ldots, w_{n+})$ and $w_{i+} = \sum_j w_{ij}$ is the i^{th} row sum of the binary adjacency matrix $\mathbf{W} = (w_{ij})$, 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 follows

$$\gamma \sim N(\mathbf{0}, [\tau_\gamma \mathbf{R}_\gamma]^-),$$

where τ_γ is a precision parameter and $\mathbf{R}_\gamma$ is the $T \times T$ structure matrix of a RW1/RW2 (see [12], 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

$$\delta \sim N(\mathbf{0}, [\tau_\delta \mathbf{R}_\delta]^-),$$

where τ_δ is a precision parameter and $\mathbf{R}_\delta$ 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 were originally proposed by Knorr-Held [26] (see Table 1).

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

Interaction	$\mathbf{R}_\delta$	Spatial correlation	Temporal correlation
Type I	$\mathbf{I}_T \otimes \mathbf{I}_n$	-	-
Type II	$\mathbf{R}_\gamma \otimes \mathbf{I}_n$	-	✓
Type III	$\mathbf{I}_T \otimes \mathbf{R}_\xi$	✓	-
Type IV	$\mathbf{R}_\gamma \otimes \mathbf{R}_\xi$	✓	✓

Table 2
Identifiability constraints for the different types of space-time interaction effects in CAR models [31].

Interaction	$\mathbf{R}_\delta$	Constraints
Type I	$\mathbf{I}_T \otimes \mathbf{I}_n$	$\sum_{i=1}^n \xi_i = 0$, $\sum_{t=1}^T \gamma_t = 0$, and $\sum_{i=1}^n \sum_{t=1}^T \delta_{it} = 0$
Type II	$\mathbf{R}_\gamma \otimes \mathbf{I}_n$	$\sum_{i=1}^n \xi_i = 0$, $\sum_{t=1}^T \gamma_t = 0$, and $\sum_{t=1}^T \delta_{it} = 0$, for $i = 1, \ldots, n$
Type III	$\mathbf{I}_T \otimes \mathbf{R}_\xi$	$\sum_{i=1}^n \xi_i = 0$, $\sum_{t=1}^T \gamma_t = 0$, and $\sum_{i=1}^n \delta_{it} = 0$, for $t = 1, \ldots, T$
Type IV	$\mathbf{R}_\gamma \otimes \mathbf{R}_\xi$	$\sum_{i=1}^n \xi_i = 0$, $\sum_{t=1}^T \gamma_t = 0$, and $\sum_{t=1}^T \delta_{it} = 0$, for $i = 1, \ldots, n$, $\sum_{i=1}^n \delta_{it} = 0$, for $t = 1, \ldots, T$.

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 [31]. 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 [32]. Although these simulation-based techniques are widely used, mainly due to the development of free software to run MCMC algorithms such as WinBUGS [33], JAGS [34], STAN [35] or NIMBLE [36], these methods tend to be computationally very demanding and large Monte Carlo errors are usually present for complex spatio-temporal models [37]. An alternative method to improve the speed of these calculations is to approximate the marginal posteriors of the model parameters using integrated nested Laplace approximations (INLA) [38]. The INLA technique is especially attractive for latent GMRFs with sparse precision matrices and is being increasingly used in applied statistics in general [39] and in the field of spatial statistics in particular [40]. Recently, NIMBLE and R-INLA have been compared in a simulation study to fit spatio-temporal disease mapping models [41]. The results obtained are identical in terms of relative risk estimates and nearly identical in terms of parameter estimates. However, R-INLA is considerably faster than NIMBLE.

3. Methodology

There is no doubt that the use of spatio-temporal CAR models allows accurate risk estimates to be obtained in reasonable computational times when the number of areal-units is relatively small.

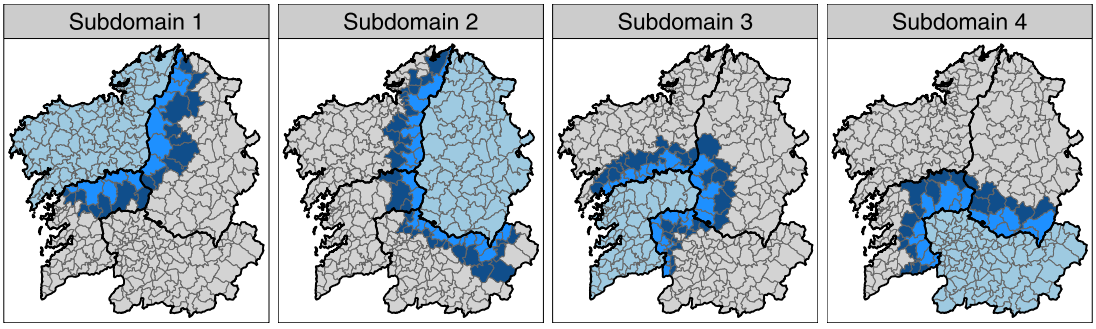

Fig. 1. Toy example of a spatial partition into $D = 4$ subdomains. Light-blue areas represent those corresponding to the Disjoint models, while spatial adjacent areas are added when considering the 1st/2nd-order neighbourhood models (blue and dark-blue areas, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 observations (spatial locations $\times$ time points), leading to prohibitive computational times if (partial) matrix inversions are necessary during the estimation process. In addition, a CAR model's spatial-dependence parameter is constant throughout 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 spatial models described in Orozco-Acosta et al. [25] 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 Eq. (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. [25], *Disjoint* and *k-order neighbourhood models* are defined for estimating spatio-temporal disease risks.

For the *Disjoint model*, a partition of the spatial domain $\mathscr{D}$ into D subdomains is defined, so that $\mathscr{D} = \bigcup_{d=1}^{D} \mathscr{D}_d$ where $\mathscr{D}_j \cap \mathscr{D}_k = \emptyset$ for all $j \neq k$. If we denote as A_{it} the small area i in time period t, let $\mathbf{O}_d = \{O_{it} \,|\, A_{it} \in \mathscr{D}_d\}$ and $\mathbf{E}_d = \{E_{it} \,|\, A_{it} \in \mathscr{D}_d\}$ 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 $\log \mathbf{r} = (\log \mathbf{r}_1, \ldots, \log \mathbf{r}_D)'$ 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. Note that under this modelling proposal, some A_{it} units located at the borders of the partitions will be included in different submodels. That is, $\sum_{d=1}^{D} n_d > nT$, where n_d denotes the number of observations within the partition $\mathscr{D}_d$. In consequence, the final log-risk surface is no longer the union of the posterior estimates of the relative risks r_{it} obtained from each submodel as some areal-time units would have more than one estimated posterior distribution. Two different merging strategies can be considered to properly combine their posterior estimates. Originally, [25] proposed to compute mixture distributions of the estimated posterior probability density functions with weights proportional to the conditional predictive ordinates (CPOs[42];). 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 5.

4. Software: Model implementation in the R package `bigDM`

We have implemented several scalable spatio-temporal disease mapping models in the R package `bigDM` [43] (see https://github.com/spatialstatisticsupna/bigDM). A generic scheme of the main algorithms is 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, we consider only purely spatial partitions. 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 estimates.

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 [44]. 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 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 combine 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

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 (ξ), temporal (γ) and spatio-temporal (δ) 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

if `W=NULL` **then**
  compute `W` from the cartography file.
Merge disjoint connected subgraphs.
Define `formula` object for INLA model according to the prior distributions for ξ, γ and δ.

Step 2: Fitting submodels with INLA

Divide the spatial domain into D subdomains.
for $d \in \{1, \ldots, D\}$ **do**
  if `k>0` **then**
    add k-order neighbouring areas.
  Compute the spatial adjacency matrix W_d.
  Extract $\mathbf{O}_d = \{O_{it} | A_{it} \in \mathscr{D}_d\}$ and $\mathbf{E}_d = \{E_{it} | A_{it} \in \mathscr{D}_d\}$.
  Define appropriate identifiability constraints.
  if `plan="sequential"` **then**
    fit INLA models sequentially in the current R session (local machine).
  else
    fit INLA models in parallel on external R sessions (local machine) or distributed in remote compute nodes.

Step 3: Merging results

if `plan="cluster"` **then**
  retrieve submodels to the central node.
if `k>0` and `merge.strategy="mixture"` **then**
  compute mixture distributions for the posterior probability density functions of each $\log r_{it}$. ▷ Algorithm 2
if `k>0` and `merge.strategy="original"` **then**
  select the posterior marginal estimate of the areal-unit corresponding to the original subdomain
Compute approximate DIC and WAIC values. ▷ Algorithm 3

Output:
- `inla` object with the fitted model.

computed for each $\log r_{it}$ (see Algorithm 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) [45] and the Watanabe-Akaike information criterion (WAIC) [46], 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. Details about the computations of these approximations are given in Algorithm 3 (see [47] for further details). Specific vignettes ac-

Algorithm 2 Compute mixture distributions for each $\log r_{it}$.

Inputs: `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}.
- $CPO_{it}^d = Pr(O_{it} = o_{it} | \mathbf{o}_{-it})$: 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

for $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$ **do**
  Compute $m(i)$: number of submodels in which the areal-unit A_{it} has been estimated (note that $m(i) < D$)
  Compute normalized weights

$$w_j = \frac{CPO_{it}^j}{\sum_j CPO_{it}^j}, \quad \text{for} \quad j = 1, \ldots, m(i).$$

  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.

companying the package have been included to facilitate the use of the methodology for non-expert users. See https://emi-sstcdapp.unavarra.es/bigDM/bigDM-3-fitting-spatio-temporal-models.html.

5. 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 computing speed offered by our modelling approach when using both parallel and/or distributed computation strategies as the number of small areas increases.

5.1. Risk estimation in high-dimensional areal data

5.1.1. Data generation

The $n = 7907$ municipalities of continental Spain and $T = 25$ time periods are used as the simulation template. Under this template, we generate a smooth risk surface 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. 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. 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} is fixed at value 10 and r_{it} are the true generated risks. We generate a total of 50 simulated datasets.

5.1.2. Results

We fitted four different spatio-temporal models to each simulated dataset: the Global model (Section 2) and the Disjoint, 1st-order neighbourhood and 2nd-order neighbourhood scalable model

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** $i \in \{1,\dots,n\}$ and $t \in \{1,\dots,T\}$ **do**

2: Draw S samples from the posterior marginal distribution of r_{it}.

3: Compute θ^s: posterior simulations of $\mu_{it} = E_{it} r_{it}$.

4: Compute the deviance information criterion $\text{DIC} = 2\overline{D(\theta)} - D(\bar{\theta})$, by approximating the mean deviance $\overline{D(\theta)}$ and the deviance of the mean $D(\bar{\theta})$ as

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

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

where $p(\mathbf{O}|\theta)$ is the likelihood function of a Poisson distribution with mean θ.

5: Approximate Watanabe-Akaike information criterion as

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

Output:

- Approximate DIC and WAIC values.

proposals (Section 3). For all these models, we consider 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 [26] for the spatio-temporal random effect.

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 `bigDM` package.

We evaluate the model performance in terms of how well the 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

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

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

where r_{it} is the true generated risk, and $\hat{r}_{it}^l$ 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., [48]) 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 the `''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 ($\approx 4 \times 10^{10}$ elements) and the high number of identifiability constraints over the spatio-temporal interaction (≈ 8000 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 (recall 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 other hand, 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 is shown in Table 3, the `''original''` merging strategy is less computationally demanding than using mixture distributions.

Scalable 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). At the bottom of Figure 2 we show the posterior median estimates of relative risks for a randomly selected simulation obtained with 2nd-order neighbourhood and Type IV interaction model using the `''original''` merging strategy. As can be seen, the values obtained are quite similar to those of the actual

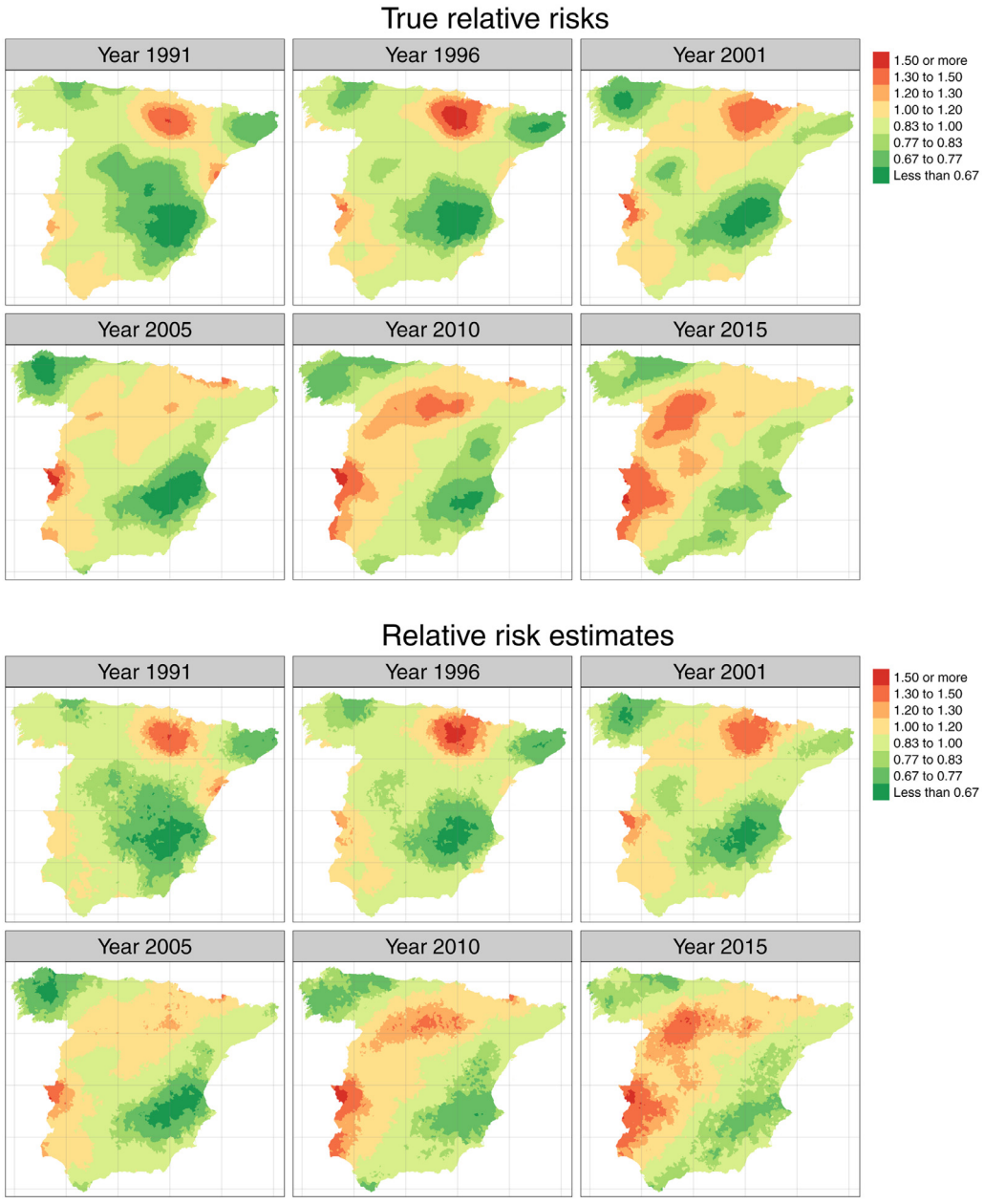

Fig. 2. Simulation study: true risk surfaces (top) and posterior median estimates of relative risks for a randomly selected simulation obtained with 2nd-order neighbourhood and Type IV interaction model using the ''original'' merging strategy (bottom) for some selected years.

risk surface. Average values of posterior median estimates of relative risks over the 50 simulated datasets are shown in Figure A.5.

In summary, we remark that in terms of model selection criteria, risk estimation accuracy and computational time, our simulation study shows that the scalable 2nd-order neighbourhood model clearly outperforms the classical Global models used in space-time disease mapping.

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 probability that r_{it} exceeds 1 is higher than a threshold value p_0, namely the exceedence probabilities $P(r_{it} > 1|\mathbf{O}) > p_0$. Conversely, we classify a low risk area if the posterior probabilities that r_{it} is below 1 is higher than p_0, i.e., $P(r_{it} < 1|\mathbf{O}) > p_0$. 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 note that our proposed scalable models outperform the Global models in terms of high and low risk area detection. In particular, the first order neighborhood model ''original'' strategy (Type IV interaction) performs the best in terms of TPR. The rest of the

Table 3
Simulation study: average values of mean deviance $\overline{D(\theta)}$, 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 selection criteria				Risk estimation accuracy			Time	
Model	Interaction	$\overline{D(\theta)}$	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"										
1st order neighbourhood	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	0.2490	10	18
	Type IV	198603	5243	203846	203823	0.0133	0.0311	0.1974	70	18
2nd order neighbourhood	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	0.2593	16	32
	Type IV	198477	5397	203874	203848	0.0134	0.0312	0.2008	136	32
merge.strategy="original"										
1st order neighbourhood	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	0.2395	10	8
	Type IV	198976	4885	203861	203846	0.0137	0.0312	0.1911	70	8
2nd order neighbourhood	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

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 Table 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.

In conclusion, our simulation study shows that, even when generating the true risk surfaces using a model that differs from those used to analyze the data, our scalable model proposals outperform the classical Global models in terms of risk estimation accuracy, detection of high and low risk areas and avoidance of false alarms. For practitioners, we recommend the use of k-order neighbourhood models with an `''original''` merging strategy.

5.2. 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 Eq. (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 4×4 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 Fig. 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.

6. 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 [49], 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 19.5% of cancer mortality [50]. 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 [51].

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 $P(r_{it} > 1|\mathbf{O})$ and $P(r_{it} < 1|\mathbf{O})$, respectively. For the 1st/2nd-order neighbourhood models, both `''mixture''` and `''original''` merging strategies are compared.

		True Positive Rate			True Negative Rate			False Positive Rate			False Negative Rate		
Model	Space-time interaction	$p_0 = 0.8$	$p_0 = 0.9$	$p_0 = 0.95$	$p_0 = 0.8$	$p_0 = 0.9$	$p_0 = 0.95$	$p_0 = 0.8$	$p_0 = 0.9$	$p_0 = 0.95$	$p_0 = 0.8$	$p_0 = 0.9$	$p_0 = 0.95$
Global	Type I	0.5932	0.4349	0.3212	0.6902	0.5453	0.4249	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 neighbourhood	Type I	0.7685	0.6601	0.5688	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 neighbourhood	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 neighbourhood	Type I	0.7736	0.6698	0.5832	0.8406	0.7557	0.6699	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	0.7532	0.6699	0.0034	0.0006	0.0001	0.0063	0.0013	0.0003
	Type IV	**0.8260**	**0.7461**	**0.6782**	**0.8798**	**0.8117**	**0.7462**	0.0037	0.0008	0.0002	0.0064	0.0014	0.0003
2nd order neighbourhood	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	0.7501	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

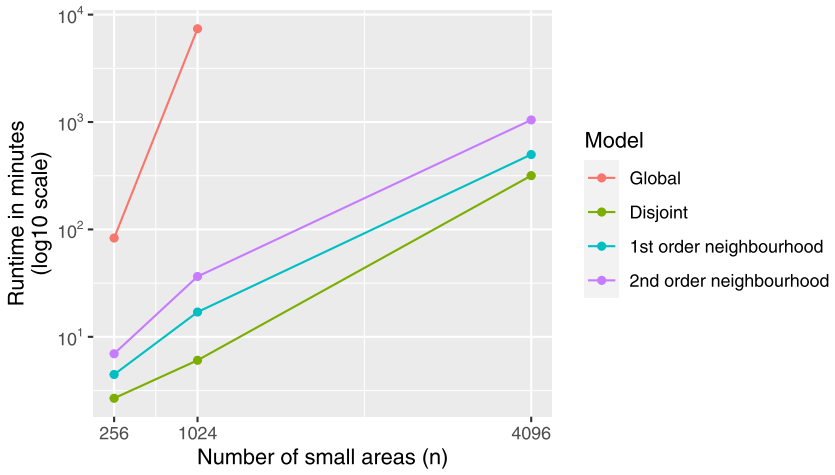

Fig. 3. Numerical simulation: computational time (in log10 scale) vs number of small areas for the Global model and our scalable modelling proposals.

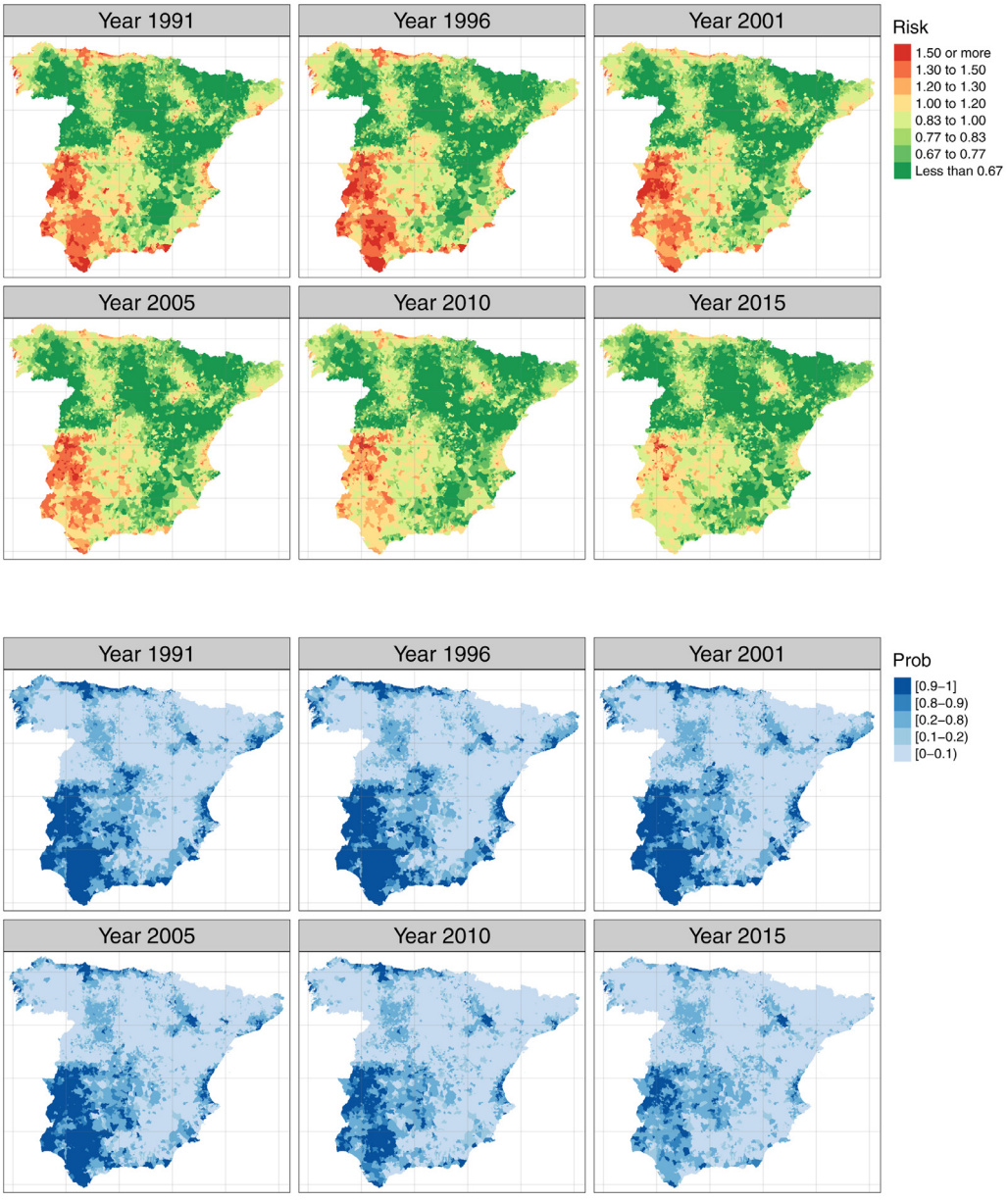

Fig. 4. Maps of posterior median estimates of relative risks r_{it} (top) and posterior exceedence probabilities $P(r_{it} > 1|\mathbf{O})$ (bottom) for the 1st-order neighbourhood model considering a BYM2 conditional autoregressive prior for space, RW1 prior for time and Type IV interaction for the spatio-temporal effect.

Table 5
Lung cancer observed deaths, expected deaths and standardized mortality ratios (SMR) for the provincial capital municipalities during the period 1991–2015.

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	2693	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	357.2	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	$\overline{D(\theta)}$	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 neighbourhood	Type I	143269	3824	147094	147112	14	8	22
	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 neighbourhood	Type I	143500	3603	147103	147138	19	10	28
	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

$\overline{D(\theta)}$: 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

A total of 378,720 lung cancer deaths (corresponding to International Classification of Diseases-10 codes C33-C34) were registered for the 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 was computed using the indirect (internal) standardization method with 5-year 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 distributed 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 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 show slightly better performance. Maps with the posterior median estimates of relative risks and posterior exceedence probabilities $P(r_{it} > 1|\mathbf{O})$ obtained with the 1st-order neighbourhood model considering a Type IV interaction for the spatio-temporal random effect are plotted in Fig. 4. The estimated risk surfaces are consistent with those described by [52], where the geographical pattern of lung cancer mortality data in Spain at municipality level was analyzed using spatial models.

7. Conclusions

The use of spatial and spatio-temporal hierarchical models for regional data are crucial in areas such as cancer epidemiology, since they allow one to obtain reliable incidence or mortality risk estimates of cancer in small areas, avoiding the huge variability of classical risk estimation measures such as the standardized mor-

tality 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 areas 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 computing power 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 is appropriate when the number of small areas is 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. Futhermore, 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 the `''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 into 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 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 `R-INLA`, as 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 [53,54]. This new approximation technique has been shown to provide accurate results with computational efficiency and scalability superior to the classic integrated nested Laplace approximations. Recent versions of the `bigDM` package ($>= 0.5.1$) are compatible with this new avenue for Bayesian inference with INLA by including the `inla.mode=''compact''` argument in its main functions. For further details, see the reference manual of the package.

Finally, we are currently working on extending our Bayesian modelling proposal to ecological regression models that take into account the spatial and/or spatio-temporal confounding issues between fixed and random effects [55], as well as to high-dimensional multivariate disease mapping models in which several diseases are jointly analyzed [56].

Declaration of Competing Interest

The authors do not have any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

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 improve this paper. Thanks are also given to two anonymous reviewers for their careful reading of our manuscript and their insightful comments that helped to clarify it. Open access funding provided by Universidad Pública de Navarra.

Appendix A

Table A1
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**.

		merge.strategy="mixture"			merge.strategy="original"		
Model	Interaction	MARB	MRRMSE	IS	MARB	MRRMSE	IS
1st order neighbourhood	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 neighbourhood	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 A2
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 $P(r_{it} > 1|\mathbf{O})$. Results for the 1st/2nd-order neighbourhood models **computed only for the border areas**.

True Positive Rate							
		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	0.7598	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							
		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	0.0054	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 A3
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 $P(r_{it} < 1|\mathbf{O})$. Results for the 1st/2nd-order neighbourhood models **computed only for the border areas**.

True Negative Rate							
		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
False Negative Rate							
		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

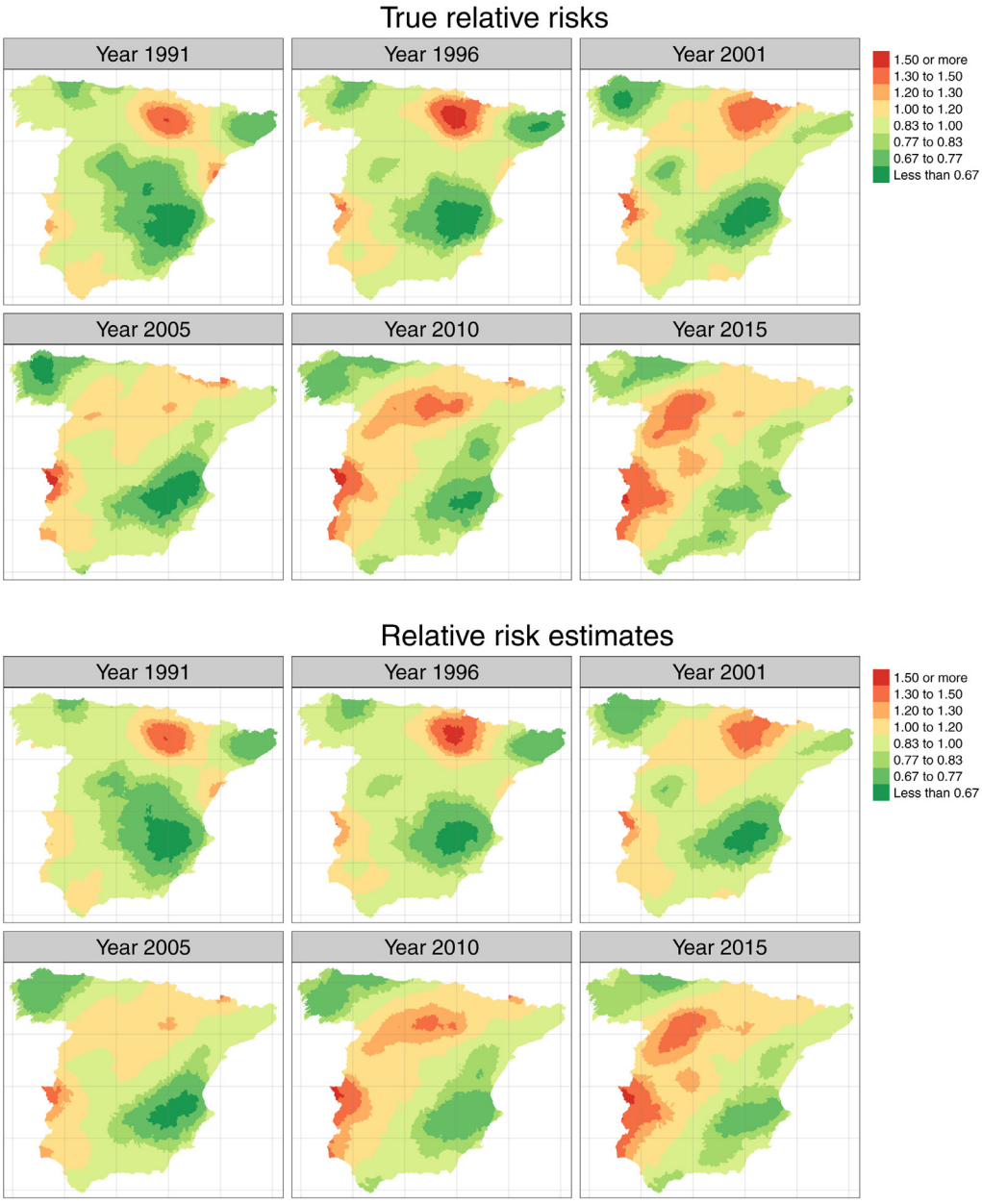

Fig. A1. Simulation study: true risk surfaces (top) and average values of posterior median estimates of relative risks for 2nd-order neighbourhood and Type IV interaction model using the `''original''` merging strategy (bottom) for some selected years.

References

[1] N. Cressie, C.K. Wikle, Statistics for Spatio-Temporal Data, Wiley Series in Probability and Statistics, John Wiley & Sons, 2011.

[2] S. Banerjee, B.P. Carlin, A.E. Gelfand, Hierarchical Modeling and Analysis for Spatial Data, Monographs on Statistics and Applied Probability, volume 135, 2nd, CRC Press, Boca Raton, FL, 2015.

[3] Y. Sun, B. Li, M.G. Genton, Geostatistics for large datasets, in: Advances and Challenges in Space-time Modelling of Natural Events, Springer, 2012, pp. 55–77.

[4] S. Banerjee, High-dimensional Bayesian geostatistics, Bayesian Analysis 12 (2) (2017) 583, doi:10.1214/17-BA1056R.

[5] M.J. Heaton, A. Datta, A.O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R.B. Gramacy, D. Hammerling, M. Katzfuss, et al., A case study competition among methods for analyzing large spatial data, Journal of Agricultural, Biological and Environmental Statistics 24 (3) (2019) 398–425, doi:10.1007/s13253-018-00348-w.

[6] H. Liu, Y.-S. Ong, X. Shen, J. Cai, When gaussian process meets big data: a review of scalable GPs, IEEE Trans Neural Netw Learn Syst 31 (11) (2020) 4405–4423, doi:10.1109/TNNLS.2019.2957109.

[7] M. Appel, E. Pebesma, Spatiotemporal multi-resolution approximations for analyzing global environmental data, Spat Stat 38 (2020) 100465, doi:10.1016/j.spasta.2020.100465.

[8] M. Katzfuss, A multi-resolution approximation for massive spatial datasets, J Am Stat Assoc 112 (517) (2017) 201–214, doi:10.1080/01621459.2015.1123632.

[9] A. Zammit-Mangion, J. Rougier, Multi-scale process modelling and distributed computation for spatial data, Stat Comput 30 (6) (2020) 1609–1627, doi:10.1007/s11222-020-09962-6.

[10] F. Lindgren, H. Rue, J. Lindström, 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) (2011) 423–498, doi:10.1111/j.1467-9868.2011.00777.x.

[11] A.V. Vecchia, Estimation and model identification for continuous spatial processes, Journal of the Royal Statistical Society: Series B (Methodological) 50 (2) (1988) 297–312, doi:10.1111/j.2517-6161.1988.tb01729.x.
[12] H. Rue, L. Held, Gaussian Markov random fields: Theory and Applications, CRC Press, 2005.
[13] A.O. Finley, A. Datta, B.D. Cook, D.C. Morton, H.E. Andersen, S. Banerjee, Efficient algorithms for Bayesian nearest neighbor Gaussian processes, Journal of Computational and Graphical Statistics 28 (2) (2019) 401–414, doi:10.1080/10618600.2018.1537924.
[14] A. Datta, S. Banerjee, A.O. Finley, A.E. Gelfand, Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets, J Am Stat Assoc 111 (514) (2016) 800–812, doi:10.1080/01621459.2015.1044091.
[15] M. Peruzzi, S. Banerjee, A.O. Finley, Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains, J Am Stat Assoc (2020) 1–14, doi:10.1080/01621459.2020.1833889.
[16] M. Katzfuss, J. Guinness, A general framework for vecchia approximations of Gaussian processes, Statistical Science 36 (1) (2021) 124–141, doi:10.1214/19-STS755.
[17] M. Jurek, M. Katzfuss, Hierarchical sparse cholesky decomposition with applications to high-dimensional spatio-temporal filtering, Stat Comput 32 (1) (2022) 1–19, doi:10.1007/s11222-021-10077-9.
[18] A.B. Lawson, S. Banerjee, R.P. Haining, M.D. Ugarte, Handbook of Spatial Epidemiology, CRC Press, 2016.
[19] W. Shen, T.A. Louis, Triple-goal estimates for disease mapping, Stat Med 19 (17–18) (2000) 2295–2308.
[20] Y. Guan, M. Haran, A computationally efficient projection-based approach for spatial generalized linear mixed models, Journal of Computational and Graphical Statistics 27 (4) (2018) 701–714, doi:10.1080/10618600.2018.1425625.
[21] J.S. Hodges, B.J. Reich, Adding spatially-correlated errors can mess up the fixed effect you love, Am Stat 64 (4) (2010) 325–334, doi:10.1198/tast.2010.10052.
[22] J. Hughes, M. Haran, Dimension reduction and alleviation of confounding for spatial generalized linear mixed models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75 (1) (2013) 139–159, doi:10.1111/j.1467-9868.2012.01041.x.
[23] A. Datta, S. Banerjee, J.S. Hodges, L. Gao, Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models, Bayesian Analysis 14 (4) (2019) 1221, doi:10.1214/19-ba1177.
[24] L. Gao, A. Datta, S. Banerjee, Hierarchical multivariate directed acyclic graph autoregressive models for spatial diseases mapping, Stat Med 41 (16) (2022) 3057–3075, doi:10.1002/sim.9404.
[25] E. Orozco-Acosta, A. Adin, M.D. Ugarte, Scalable Bayesian modelling for smoothing disease risks in large spatial data sets using INLA, Spat Stat 41 (2021) 100496, doi:10.1016/j.spasta.2021.100496.
[26] L. Knorr-Held, Bayesian modelling of inseparable space-time variation in disease risk, Stat Med 19 (17–18) (2000) 2555–2567, doi:10.1002/1097-0258(20000915/30)19:17/18⟨2555::AID-SIM587⟩3.0.CO;2-#.
[27] C.B. Dean, M.D. Ugarte, A.F. Militino, Detecting interaction between random regions and fixed age effects in disease mapping, Biometrics 57 (1) (2001) 197–202, doi:10.1111/j.0006-341X.2001.00197.x.
[28] A. Riebler, S.H. Sørbye, D. Simpson, H. Rue, An intuitive Bayesian spatial model for disease mapping that accounts for scaling, Stat Methods Med Res 25 (4) (2016) 1145–1165, doi:10.1177/0962280216660421.
[29] B.G. Leroux, X. Lei, N. Breslow, Estimation of disease rates in small areas: a new mixed model for spatial dependence, in: M. Halloran, D. Berry (Eds.), Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer-Verlag: New York, 1999, pp. 179–191.
[30] S.H. Sørbye, H. Rue, Scaling intrinsic Gaussian Markov random field priors in spatial modelling, Spat Stat 8 (2014) 39–51, doi:10.1016/j.spasta.2013.06.004.
[31] T. Goicoa, A. Adin, M.D. Ugarte, J.S. Hodges, In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results, Stochastic Environmental Research and Risk Assessment 32 (3) (2018) 749–770, doi:10.1007/s00477-017-1405-0.
[32] W.R. Gilks, S. Richardson, D. Spiegelhalter, Markov chain Monte Carlo in practice, CRC press, 1995.
[33] D. Spiegelhalter, A. Thomas, N. Best, D. Lunn, WinBUGS user manual, 2003.
[34] M. Plummer, et al., 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, Vienna, Austria., 2003, pp. 1–10.
[35] B. Carpenter, A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell, Stan: a probabilistic programming language, J Stat Softw 76 (1) (2017) 1–32.
[36] P. de Valpine, D. Turek, C.J. Paciorek, C. Anderson-Bergman, D.T. Lang, R. Bodik, Programming with models: writing statistical algorithms for general model structures with NIMBLE, Journal of Computational and Graphical Statistics 26 (2) (2017) 403–413, doi:10.1080/10618600.2016.1172487.
[37] B. Schrödle, L. Held, A. Riebler, J. Danuser, 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) (2011) 261–279, doi:10.1111/j.1467-9876.2010.00740.x.
[38] H. Rue, S. Martino, N. Chopin, 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) (2009) 319–392, doi:10.1111/j.1467-9868.2008.00700.x.
[39] H. Rue, A. Riebler, S.H. Sørbye, J.B. Illian, D.P. Simpson, F.K. Lindgren, Bayesian computing with INLA: a review, Annu Rev Stat Appl 4 (2017) 395–421, doi:10.1146/annurev-statistics-060116-054045.
[40] H. Bakka, H. Rue, G.-A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, F. Lindgren, Spatial modeling with R-INLA: a review, Wiley Interdiscip. Rev. Comput. Stat. 10 (6) (2018) e1443, doi:10.1002/wics.1443.
[41] A. Urdangarin, T. Goicoa, M.D. Ugarte, Space-time interactions in Bayesian disease mapping with recent tools: making things easier for practitioners, Stat Methods Med Res 31 (6) (2022) 1085–1103, doi:10.1177/09622802221079351.
[42] L. Pettit, The conditional predictive ordinate for the normal distribution, Journal of the Royal Statistical Society: Series B (Methodological) 52 (1) (1990) 175–184, doi:10.1111/j.2517-6161.1990.tb01780.x.
[43] A. Adin, E. Orozco-Acosta, M.D. Ugarte, bigDM: Scalable Bayesian Disease Mapping Models for High-Dimensional Data, 2022. R package version 0.4.2, https://github.com/spatialstatisticsupna/bigDM.
[44] H. Bengtsson, A unifying framework for parallel and distributed processing in R using futures, R J 13 (2) (2021) 273–291, doi:10.32614/RJ-2021-048.
[45] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4) (2002) 583–639, doi:10.1111/1467-9868.00353.
[46] S. Watanabe, Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, Journal of Machine Learning Research 11 (Dec) (2010) 3571–3594.
[47] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models, Stat Comput 24 (6) (2014) 997–1016, doi:10.1007/s11222-013-9416-2.
[48] T. Gneiting, A.E. Raftery, Strictly proper scoring rules, prediction, and estimation, J Am Stat Assoc 102 (477) (2007) 359–378, doi:10.1198/016214506000001437.
[49] J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, F. Bray, Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018, Eur J Cancer 103 (2018) 356–387, doi:10.1016/j.ejca.2018.07.005.
[50] Spanish Society of Medical Oncology (Sociedad Española de Oncología Médica), Las cifras del cáncer en España 2020, 2020. https://seom.org/seomcms/images/stories/recursos/Cifras_del_cancer_2020.pdf.
[51] J. Remon, et al., Lung cancer in Spain, Journal of Thoracic Oncology 16 (2) (2021) 197–204, doi:10.1016/j.jtho.2020.09.026.
[52] G. López-Abente, N. Aragonés, B. Pérez-Gómez, M. Pollán, J. García-Pérez, R. Ramis, P. Fernández-Navarro, Time trends in municipal distribution patterns of cancer mortality in spain, BMC Cancer 14 (1) (2014) 1–15, doi:10.1186/1471-2407-14-535.
[53] J. van Niekerk, H. Rue, Correcting the Laplace method with variational bayes, arXiv preprint (2021), doi:10.48550/arXiv.2111.12945.
[54] J. van Niekerk, E. Krainski, D. Rustand, H. Rue, A new avenue for Bayesian inference with INLA, Computational Statistics & Data Analysis 107692 (2023), doi:10.1016/j.csda.2023.107692.
[55] A. Adin, T. Goicoa, J.S. Hodges, P.M. Schnell, M.D. Ugarte, Alleviating confounding in spatio-temporal areal models with an application on crimes against women in india, Statistical Modelling (published online on May 31, 2021) (2021), doi:10.1177/1471082X211015452.
[56] G. Vicente, T. Goicoa, M.D. Ugarte, Multivariate Bayesian spatio-temporal P-spline models to analyze crimes against women, Biostatistics (published online on Dec 27, 2021) (2021), doi:10.1093/biostatistics/kxab042.