Flexible Bayesian P-splines for smoothing age-specific spatio-temporal mortality patterns

Goicoa T[1,2,3,*], Adin A[1,2], Etxeberria J[1,2,4], Militino AF[1,2], and Ugarte MD[1,2]

(1)Department of Statistics and Operations Research, Public University of Navarre, Spain

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

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

(4)Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Spain

∗Correspondence to Tomás Goicoa, Department of Statistics and Operations Research, Public University of Navarre, Campus de Arrosadía, 31006 Pamplona, Spain.

E-mail: tomas.goicoa@unavarra.es

Abstract

In this paper age-space-time models based on one and two-dimensional P-splines with B-spline bases are proposed for smoothing mortality rates, where both fixed relative scale and scale invariant two-dimensional penalties are examined. Model fitting and inference are carried out using integrated nested Laplace approximations (INLA), a recent Bayesian technique that speeds up computations compared to McMC methods. The models will be illustrated with Spanish breast cancer mortality data during the period 1985-2010, where a general decline in breast cancer mortality has been observed in Spanish provinces in the last decades. The results reveal that mortality rates for the oldest age groups do not decrease in all provinces.

Keywords: Breast cancer mortality; disease mapping; INLA; smoothing; time-space-age models.

1 Introduction

Splines and penalized splines have been recently consolidated into the disease mapping toolkit as an alternative to the widely used conditional autoregressive (CAR) models[1] for smoothing and identifying the overall geographical pattern of risks and/or rates. P-splines have been popularized by Eilers and Marx,[2] who consider B-spline bases and penalties based on finite order differences of adjacent coefficients. P-splines also have their Bayesian counterparts replacing the penalties on the coefficients by their stochastic analogs, that is, using a random walk of first or second order as a prior distribution on the coefficients (see for example Lang and Brezger[3]).
Spatio-temporal models incorporating splines have been used in space-time disease mapping either from an empirical Bayes (EB) or a fully Bayes (FB) perspective. Within the EB setting, B-splines were first used to model temporal components whereas CAR models were adopted for the spatial effects.[4] Later on, Ugarte et al.[5,6] use three-dimensional P-spline models to account for the spatial and the temporal effects. These authors reformulate the P-spline models as mixed models and used penalized quasi-likelihood (PQL)[7] for model fitting and inference. From a FB approach, MacNab and Gustafson[8] consider unpenalized regression B-splines to illustrate

how region-specific temporal relative risk trends may evolve in either a spatially structured or spatially unstructured manner. MacNab[9] compares unpenalized regression B-splines, smoothing splines, and P-splines to analyze the connections and differences among them, and also evaluates their capabilities to smooth risk trends in a disease mapping context. The author considers the mixed model reformulation of P-splines. This is not the approach followed by Bauer et al.[10] consisting in using random walks of different order as prior distributions on the coefficients. All these models are very complex and closed form expressions for the posterior distributions of the quantities of interest are intractable. Consequently Markov chain Monte Carlo (McMC) methods are a common tool for model fitting and inference, yet these methods also present some inconveniences. For example, algorithms could be tricky and time-consuming when complex models are handled.[11]. To avoid these problems, we will adopt a new technique based on integrated nested Laplace approximations (INLA).[12]

Studying age-specific spatio-temporal mortality patterns has not been so common in the literature, yet there is some research including the age groups in the models (see for example Sun et al.[13] and Zhang et al.[14]). Age-period-cohort models have also been used to study age-specific spatio-temporal patterns. For example Papoila et al.[15] analyze female and male stomach cancer incidence data in Sourthern Portugal between 1998 and 2006, and Etxeberria et al.[16] conduct a similar study for pancreatic cancer mortality data in Spain during the period 1990-2013. Very recently, Goicoa et al.[17] propose age-specific spatio-temporal CAR models to analyze prostate cancer mortality in Spain, revealing differences in temporal mortality trends by age groups and provinces.

The goal of this paper is to propose several age-space-time P-spline models to analyze the geographical pattern of a disease and the temporal evolution of the mortality rates according to the age groups. We consider a CAR model for the spatial random effects, and P-splines for time and age. In particular, one and two-dimensional P-splines will be adopted to capture space-time, space-age, time-age, and age-space-time interactions.

The reminder of the paper is outlined as follows. Section 2 provides some background on breast cancer mortality and describes the data at hand. Section 3 introduces time-space-age P-spline models. Section 4 reviews some aspects of the INLA methodology. In Section 5 we analyze female breast cancer mortality data in Spanish provinces. Finally, Section 6 provides some discussion and conclusions.

2 Female breast cancer mortality

Breast cancer mortality corresponds to code C50 of the International Classification of Diseases-10.[18] According to the last cancer mortality estimates for 40 European countries in 2012,[19] breast cancer was the leading cause of death by cancer in females in Europe (131,000 deaths, 16.8% of total cancer deaths in females), followed by colorectal (102,000 deaths, 13.0%) and lung cancer (almost 100,000 deaths, 12.7%). It is also known that breast cancer death rates generally increase with age.

Temporal trends of female breast cancer mortality rates have been widely analyzed in Europe and particularly in Spain. In most developed countries, breast cancer mortality rates increased from the 1950s to the 1980s, and from the 1990s onward they turned around mainly due to the implementation of screening programmes and improvements in breast cancer care through new effective treatments.[20,21] In 1990, Navarre was the first Spanish region implementing a

population-based breast cancer screening programme, followed subsequently by other regions. Nowadays all Spanish screening programmes include women aged 50-69 years in their target population, although 7 out of 19 regions start at the age of 45.[22] Analyses aggregating counts over age-groups have revealed a decrease in rates after the 1990s, but some authors[23] suggest that this decline is not clear for all age groups in Spain. In particular, Ugarte et al.[24] split the female population into three age groups and use a spatio-temporal P-spline model to run three independent analyses, one for each age group. They observe that the decline in mortality is delayed for the oldest age group. However, this analysis does not allow to "borrow strength" from similar age groups and health professionals are interested in a finer division of age groups. Consequently, the use of age-space-time models provides a better knowledge about the evolution of breast cancer mortality.

Our interest here lies in estimating female breast cancer mortality rates by province, year, and age group to unveil mortality patterns in the 50 Spanish provinces during the period 1985-2010. In Spain, 96.21% of breast cancer deaths occurred in 40 years old or older women, therefore and according to experts, the following eleven age-groups are considered: $[0, 40)$, $[40, 45)$, $[45, 50)$, $[50, 55)$, $[55, 60)$, $[60, 65)$, $[65, 70)$, $[70, 75)$, $[75, 80)$, $[80, 85)$, $[+85)$. Data on population and deaths were obtained from records of the Spanish Statistical Institute. A total of 145,991 deaths were identified from 1985 to 2010. A brief summary with the number of deaths and the crude rate (number of deaths/population at risk by 100,000 female inhabitants) by age-group, province, and year is displayed in Table 1. As expected, crude rates increase with the age group. They also increase with time until late 1990s, and from then on, rates decrease slightly remaining fairly stable in the last years. Differences in crude rates among provinces seem to exist with a leap of about 17 points between the provinces with the smallest and the highest crude rates. Crude rates per 100,000 inhabitants by age, province, and year range from 0 to 389.4, with a mean of 47.61, a median of 58.92, and a standard deviation of 50.88.

Table 1: Breast cancer deaths and crude rates (by 100,000 female inhabitants) by age-group, province, and year.

age-group	cases	crude rate	province	cases	crude rate	year	cases	crude rate
$[-40)$	5530	1.940	Jaén	1660	19.568	1985	4273	21.905
$[40, 45)$	6644	18.130	Almería	1373	20.241	1986	4547	23.234
$[45, 50)$	9357	27.971	Granada	2262	20.718	1987	4880	24.868
$[50, 55)$	11875	38.083	Córdoba	2298	22.688	1988	5118	26.016
$[55, 60)$	14035	47.166	Tenerife	2482	23.070	1989	5268	26.717
$[60, 65)$	15073	53.118	...	...	...	...	...	...
$[65, 70)$	15788	60.103	Teruel	590	32.372	2006	5927	26.610
$[70, 75)$	16556	69.656	Lleida	1575	32.681	2007	5959	26.311
$[75, 80)$	16992	85.945	Asturias	4867	33.412	2008	6016	26.145
$[80, 85)$	15310	109.345	Zaragoza	3853	33.502	2009	6089	26.260
$[85+)$	18831	165.289	Huesca	980	36.083	2010	6264	26.883

Figure 1 displays the temporal evolution of crude rates in Spain. The left plot revels a general decline of mortality in the youngest age groups in Spain, but the decrease is not clear in the oldest age groups (right plot). At country level, the crude rates are not too variable, but a look at the temporal evolution of crude rates by province and age group reveals a high variability. Figure 2 displays the temporal evolution of crude rates by age group in two provinces: Madrid and

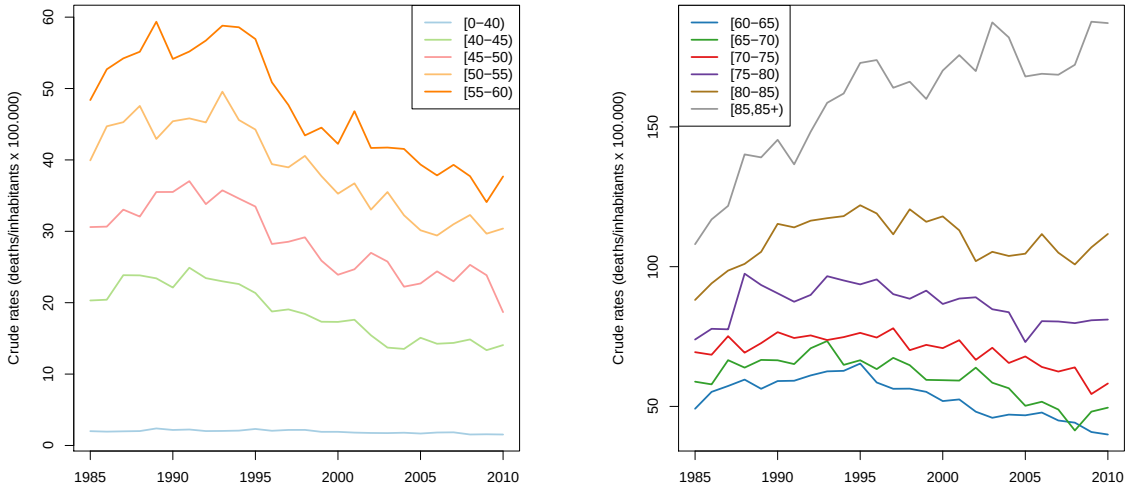

Figure 1: Temporal evolution of Spanish age-specific crude rates by 100,000 females for the eleven age groups. On the left the youngest age groups, and on the right the oldest age groups.

Huelva. Crude rates are more stable in Madrid, something expected as it is a highly populated area. It can be observed that trends decrease for the youngest age groups but the temporal trend for the oldest age group is increasing. On the other hand, Huelva is a low populated province and crude rates are highly variable making difficult the detection of any clear trend. Then, the necessity of statistical models to smooth the rates and unveil the underlying trends is beyond doubt.

3 Age-specific space-time models based on P-splines

In this section, age-space-time models based on P-splines with B-spline bases are proposed to look into geographical and temporal mortality patterns according to the age groups. The proposed models comprise simple additive models based on one-dimensional P-splines for time and age, and more complex models including two-dimensional time-age P-splines. CAR models are considered to deal with spatial heterogeneity. CAR spatial random effects produce local smoothing and they are very convenient in Spain where some policies depend on local authorities, and then inequalities in health among the different provinces are likely to exist. In addition, Ugarte et al.[25] show that a model with CAR spatial random effects outperform a model with a two-dimensional P-spline to deal with spatial heterogeneity. Here, auxiliary information other than the province, time period, and age group is not available. The different models are explained in the following. Let us consider the number of deaths, Y_{tsa}, in a specific time-period t ($t = 1, \ldots, T$), geographical area or province s ($s = 1, \ldots, S$), and age-group a ($a = 1, \ldots, A$), and let us denote by n_{tsa} the population at risk corresponding to that time, area, and age group. Then, conditional on the time-region-age specific rate r_{tsa}, the number of deaths is assumed to

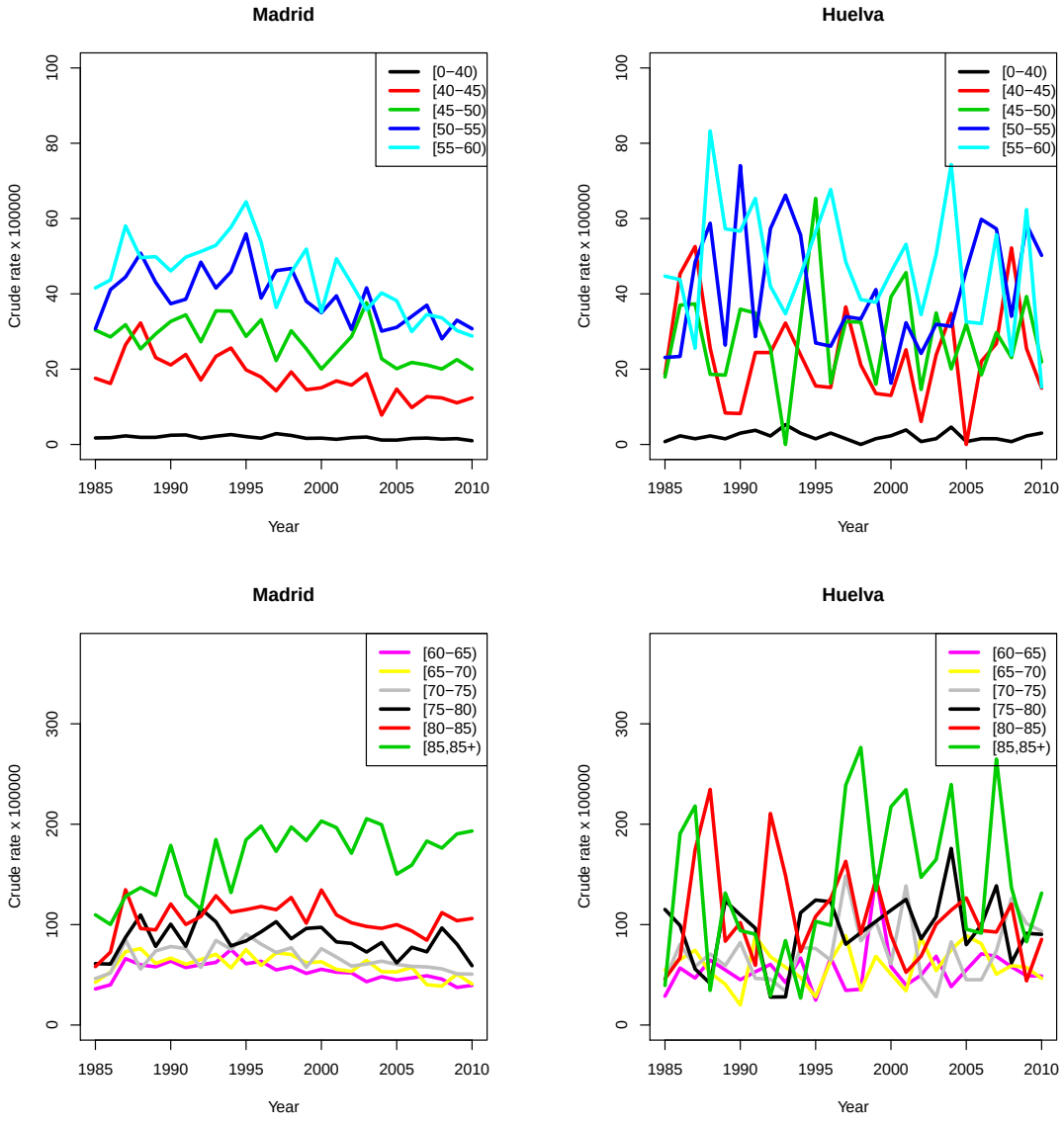

Figure 2: Temporal evolution of age-specific crude rates by 100,000 females in Madrid and Huelva for the eleven age groups. On the top row the youngest age groups, and at the bottom row the oldest age groups.

follow a Poisson distribution

$$Y_{tsa}|r_{tsa} \sim Poisson(\mu_{tsa} = n_{tsa}r_{tsa}), \quad \log \mu_{tsa} = \log n_{tsa} + \log r_{tsa}.$$

In this work several proposals based on P-splines are given to model log-rates. For simplicity and to introduce notation, let us consider the following additive P-spline model

$$\text{M1} \quad \log r_{tsa} = \gamma + \phi_s + f(x_t) + f(x_a), \tag{1}$$

where γ is an intercept, ϕ_s is the spatial random effect, and $f(x_t)$ and $f(x_a)$ are smooth functions of time (x_t) and age (x_a) respectively, that are common to all geographical areas (provinces). The spatial random effects are assumed to follow a multivariate normal distribution (see Leroux et al.[26])

$$\boldsymbol{\phi}_s \sim N(\mathbf{0}, \sigma^2_\phi \boldsymbol{\Sigma}_\phi), \quad \boldsymbol{\Sigma}_\phi = (\lambda_\phi \mathbf{Q}_\phi + (1-\lambda_\phi)\mathbf{I}_S)^{-1},$$

where $\mathbf{Q}_\phi$ is a neighborhood matrix (two areas are considered neighbors if they share a common border), $\mathbf{I}_S$ is an $S \times S$ identity matrix and λ_ϕ represents a spatial parameter taking values between 0 and 1. The prior of the spatial random effect will be denoted LCAR here after in the paper. The smooth functions can be well approximated using P-splines with B-spline bases

$$f(\mathbf{x_t}) \;\; \simeq \;\; \mathbf{B}_t\boldsymbol{\theta}_t, \;\; f(\mathbf{x_a}) \simeq \mathbf{B}_a\boldsymbol{\theta}_a,$$

where $\mathbf{B}_t$ and $\mathbf{B}_a$ are the B-spline bases for time and age respectively, $\boldsymbol{\theta}_t = (\theta_{t_1}, \ldots, \theta_{t_k})'$ and $\boldsymbol{\theta}_a = (\theta_{a_1}, \ldots, \theta_{a_m})'$ are vectors of coefficients and t_k and a_m are the number of columns of $\mathbf{B}_t$ and $\mathbf{B}_a$. Model 1 can be expressed in matrix form as

$$\text{M1} \quad \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_{SA})\boldsymbol{\theta}_t + (\mathbf{1}_{TS} \otimes \mathbf{B}_a)\boldsymbol{\theta}_a,$$

where $\mathbf{r} = (r_{111}, \ldots, r_{11A}, \ldots, r_{1S1}, \ldots, r_{1SA}, \ldots, r_{TS1} \ldots, r_{TSA})'$, $\mathbf{1}_T$, $\mathbf{1}_S$, $\mathbf{1}_A$, $\mathbf{1}_{SA}$ and $\mathbf{1}_{TS}$ represent vectors of ones of length T, S, A, $S \times A$, and $T \times S$ respectively, and $\otimes$ is the Kronecker product. First and second order random walks priors are assumed for the vector of coefficients $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_a$,[3] that is

$$\boldsymbol{\theta}_t|\lambda_t \propto exp\,(\lambda_t \boldsymbol{\theta}_t' \mathbf{P}_t \boldsymbol{\theta}_t), \qquad \boldsymbol{\theta}_a|\lambda_a \propto exp\,(\lambda_a \boldsymbol{\theta}_a' \mathbf{P}_a \boldsymbol{\theta}_a),$$

where $\mathbf{P}_t = \mathbf{D}_t'\mathbf{D}_t$, $\mathbf{P}_a = \mathbf{D}_a'\mathbf{D}_a$, and $\mathbf{D}_t$ and $\mathbf{D}_a$ are first or second order difference matrices, and λ_t and λ_a are precision parameters controlling the amount of smoothness. Prior distribution are given to the precision parameters λ_t, λ_a, and subsequent precision parameters for the different models proposed in this paper, as indicated in Section 4. The matrices $\mathbf{P}_t$ and $\mathbf{P}_a$ are just the precision matrices of a random walk of first (or second) order (see for example Rue and Held,[27] pp. 95 and 110). In Model 1, the terms $f(x_t)$ and $f(x_a)$ "borrow strength" from neighbouring time points and age groups, and the term ϕ_s is specific for each area. The spatial correlation structure allows to "borrow strength" from neighbouring areas.

Additive models may be rather restrictive in practice as interactions usually occur. That is, the temporal evolution of rates may be different in each province and age group, or the evolution of rates with age may differ among provinces or years. Hence, more flexible models including space-age, space-time, time-age, and space-time-age interactions are considered. In this paper we propose a collection of models including different types of interactions and a LCAR prior for the spatial random effect. Every model is expressed for each year, province, and age group, and in matrix form. We first start with the following interaction models using one-dimensional P-splines

$$\begin{aligned}\text{M2}\quad & \log r_{tsa} = \gamma + \phi_s + f(x_t) + f_s(x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_{SA})\boldsymbol{\theta}_t + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{sa}.\end{aligned}$$

Here, Model M2 includes a global temporal trend $f(x_t)$ common to all provinces and age groups, and province-specific age patterns $f_s(x_a)$ (space-age interactions).

$$\begin{aligned}\text{M3}\quad & \log r_{tsa} = \gamma + \phi_s + f_s(x_t) + f(x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\theta}_{st} + (\mathbf{1}_{TS} \otimes \mathbf{B}_a)\boldsymbol{\theta}_a.\end{aligned}$$

Model M3 considers an age pattern $f(x_a)$ common to all provinces and years, and province-specific temporal trends $f_s(x_t)$ (space-time interactions).

$$\begin{aligned}\text{M4}\quad & \log r_{tsa} = \gamma + \phi_s + f_s(x_t) + f_s(x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\theta}_{st} + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{sa}.\end{aligned}$$

Finally, Model M4 includes both type of interactions, province-specific temporal trends, and province-specific age patterns. The terms $f(x_t)$ and $f(x_a)$ "borrow strength" from neighbouring time points and age groups, and from all provinces as these terms are the same in the different provinces. The province-specific terms $f_s(x_t)$, and $f_s(x_a)$ "borrow strength" from neighbouring time points and age groups, respectively.
These models do not include age-time interactions, but we expect different mortality trends for all age groups (as confirmed by the exploratory data analysis). Then, five additional P-spline models are proposed to deal with age-time interactions. Some of the models combine one and two-dimensional P-splines. They are defined as

$$\begin{aligned}\text{M5}\quad & \log r_{tsa} = \gamma + \phi_s + f(x_t, x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{ta}.\end{aligned}$$

In Model M5, the term $f(x_t, x_a)$ is a smooth time-age surface common to all provinces and it is estimated using two-dimensional P-splines. It models age-time interactions allowing for different temporal trends in each age group that are common to all provinces. Through this two-dimensional P-spline, we "borrow strength" from neighbouring time points and age groups.

$$\begin{aligned}\text{M6}\quad & \log r_{tsa} = \gamma + \phi_s + f_s(x_t, x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{tsa}.\end{aligned}$$

Model M6 considers province specific two-dimensional P-splines $f_s(x_t, x_a)$ (space-time-age interaction). This term captures temporal trends that are specific for each age group and province.

$$\begin{aligned}\text{M7}\quad & \log r_{tsa} = \gamma + \phi_s + f(x_t) + f(x_a) + f_s(x_t, x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_{SA})\boldsymbol{\theta}_t + (\mathbf{1}_{TS} \otimes \mathbf{B}_a)\boldsymbol{\theta}_a \\ & \qquad + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{tsa}.\end{aligned}$$

In Model M7, we combine temporal trends and age patterns common to all provinces ($f(x_t)$ and $f(x_a)$) with time-age surfaces specific for each province ($f_s(x_t, x_a)$).

$$\begin{aligned}\text{M8}\quad & \log r_{tsa} = \gamma + \phi_s + f_s(x_t) + f_s(x_a) + f_s(x_t, x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\theta}_{st} + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{sa} \\ & \qquad + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{tsa}.\end{aligned}$$

In Model M8, smooth functions for time, age, and the time-age interaction are specific for each province. This model borrows information from neighbouring areas through the correlation structure of the spatial random effect. Information from neighbouring time points and age groups is connected through the respective smooth functions.

$$\begin{aligned} \text{M9} \quad & \log r_{tsa} = \gamma + \phi_s + f(x_t, x_a) + f_s(x_t, x_a) \\ & \log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{ta} \\ & \qquad + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{tsa}. \end{aligned}$$

Finally, Model M9 considers both a time-age surface common to all provinces and a time-age surface specific for each province. This last term can be interpreted as the specific contribution of time and age in each province.
The smooth function $f(x_t, x_a)$ can be well approximated with P-splines, that is

$$f(\mathbf{x}_t, \mathbf{x}_a) \simeq (\mathbf{B}_t \otimes \mathbf{B}_a)\boldsymbol{\theta}_{ta},$$

where $(\mathbf{B}_t \otimes \mathbf{B}_a)$ is the two-dimensional B-spline basis obtained as the Kronecker product of the marginal B-spline bases for time and age, and $\boldsymbol{\theta}_{ta}$ is the vector of coefficients. There are some details in the matrix expressions for Models M1 to M9 that deserve further comments. The terms $(\mathbf{B}_t \otimes \mathbf{1}_{SA})\boldsymbol{\theta}_t$ and $(\mathbf{1}_{TS} \otimes \mathbf{B}_a)\boldsymbol{\theta}_a$ represent the global temporal trend and the global age pattern respectively. The Kronecker product gives just a replication of the B-spline bases for time and age. The expressions $(\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\theta}_{st}$ and $(\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{sa}$ represent the area-specific temporal trend and the area-specific age pattern. Here, the vectors of coefficients $\boldsymbol{\theta}_{st}$ and $\boldsymbol{\theta}_{sa}$ include different sets of coefficients for each area but with the same smoothing parameter. Finally, $(\mathbf{B}_t \otimes \mathbf{1}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{ta}$ is the time-age surface common to all areas where $\boldsymbol{\theta}_{ta}$ is the vector of coefficients, and $(\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\,\boldsymbol{\theta}_{tsa}$ is the area-specific time-age smooth surface, where the vector $\boldsymbol{\theta}_{tsa}$ includes a set of coefficients for each area. Smoothness on the time-age two-dimensional surface is achieved by choosing appropriate prior distributions on the coefficients $\boldsymbol{\theta}_{ta}$ and $\boldsymbol{\theta}_{tsa}$. This is the Bayesian counterpart to the penalties on the coefficients in a frequentist framework. In short, the priors make close coefficients be similar. They are expressed as

$$\boldsymbol{\theta}_{ta}|\ldots \propto exp\,(\boldsymbol{\theta}'_{ta}\mathbf{P}_{ta}\boldsymbol{\theta}_{ta}), \qquad \boldsymbol{\theta}_{tsa}|\ldots \propto exp\,(\boldsymbol{\theta}'_{tsa}\mathbf{P}_{tsa}\boldsymbol{\theta}_{tsa}),$$

where $\ldots$ indicates different smoothing/precision parameters, and $\mathbf{P}_{ta}$ and $\mathbf{P}_{tsa}$ are penalty matrices taking different forms according to Table 2. These two-dimensional penalties impose smoothness on neighboring coefficients. Rearranging the coefficients by rows (time) and columns (age), the coefficient $\boldsymbol{\theta}_{ta}$ would be influenced by $\boldsymbol{\theta}_{(t-1)a}$, $\boldsymbol{\theta}_{(t+1)a}$, $\boldsymbol{\theta}_{t(a-1)}$, $\boldsymbol{\theta}_{t(a+1)}$ when we use random walks of first order. The penalties on the left have a single precision parameter λ_{ta} (top row) or λ_{tsa} (bottom row) for time and age, and the relative penalization of variability per unit change in age and time depends on the scale of the variables. We denote this penalty as "Fixed Relative Scale (FRS)". The penalties on the right have two different precision parameters for time and age, i.e, λ_{1ta}, λ_{2ta} (top row), and λ_{1tsa}, λ_{2tsa} (bottom row), and produce scale invariant smooths. They will be referred as "Scale Invariant (SI)". Scale invariant penalties are recommended when the variables are measured in different scales (see Wood et al.,[28] for an insight about scale invariant smooths).

Table 2: Two-dimensional penalties.

Fixed relative scale (FRS)	Scale invariant (SI)
$\mathbf{P}_{ta} = \lambda_{ta}(\mathbf{P}_t \otimes \mathbf{I}_{a_m} + \mathbf{I}_{t_k} \otimes \mathbf{P}_a)$	$\mathbf{P}_{ta} = (\lambda_{1ta}\mathbf{P}_t \otimes \mathbf{I}_{a_m} + \lambda_{2ta}\mathbf{I}_{t_k} \otimes \mathbf{P}_a)$
$\mathbf{P}_{tsa} = \lambda_{tsa}(\mathbf{P}_t \otimes \mathbf{I}_S \otimes \mathbf{I}_{a_m} + \mathbf{I}_{t_k} \otimes \mathbf{I}_S \otimes \mathbf{P}_a)$	$\mathbf{P}_{tsa} = (\lambda_{1tsa}\mathbf{P}_t \otimes \mathbf{I}_S \otimes \mathbf{I}_{a_m} + \lambda_{2tsa}\mathbf{I}_{t_k} \otimes \mathbf{I}_S \otimes \mathbf{P}_a)$

4 Model fitting via integrated nested Laplace approximations

McMC techniques have been the standard for Bayesian inference for the last decades. In disease mapping, they are not easy to use as models are complex and they may require large computing time when samples are highly correlated.[29] To avoid this, a new technique relying on integrated nested Laplace approximations (INLA) has been derived for latent Gaussian fields.[12] The technique is especially attractive for latent Gaussian Markov random fields with sparse precision matrices and a small number of hyperparameters because the computations are highly speeded up. The INLA approach is being increasingly used in disease mapping (see for example Schrödle and Held,[11,30] Ugarte et al.[31]), and one of its advantages is that it can be implemented in the free software `R`[32] through the R-package `R-INLA`.[33] Basically, to fit models in `R-INLA` it is necessary to define a formula object of the type

```
formula<-response~ f(.., model="",...)+....
```

where `f()` is a function `R-INLA` uses to implement the different terms of the models. These terms can be defined according to pre-specified latent Gaussian models that should be passed through the argument `model`. In this paper we consider the latent models `rw1` or `rw2` for random walks of first or second order as prior distributions for one-dimensional P-splines. They are the Bayesian counterpart to penalties of first and second order. For two-dimensional P-splines, the INLA options `generic0` and `generic3` are considered. The option `generic0` allows fitting two-dimensional P-splines with fixed relative scale penalties, and the option `generic3` is adopted to fit two-dimensional P-splines with scale invariant penalties. Full code to fit the selected model in Section 5 is available in the Appendix.

Hyperprior distributions and model selection

Prior distributions were given to all model parameters. A vague normal distribution with a precision close to zero was considered for the intercept (η). A Unif$(0, 1)$ prior was given to the spatial parameter λ_ϕ. Uniform prior distributions on the positive real line were given to the following parameters: σ_ϕ, $\sigma_t = 1/\sqrt{\lambda_t}$, $\sigma_a = 1/\sqrt{\lambda_a}$, $\sigma_{ta} = 1/\sqrt{\lambda_{ta}}$, $\sigma_{1ta} = 1/\sqrt{\lambda_{1ta}}$, $\sigma_{2ta} = 1/\sqrt{\lambda_{2ta}}$, $\sigma_{tsa} = 1/\sqrt{\lambda_{tsa}}$, $\sigma_{1tsa} = 1/\sqrt{\lambda_{1tsa}}$, and $\sigma_{2tsa} = 1/\sqrt{\lambda_{2tsa}}$. This is an improper prior and has been used in the literature providing sensible results (see for example Ugarte et al.[25] and Goicoa et al.[34]). Gelman[35] provides an in-depth discussion about prior distributions for variance parameters in hierarchical models and he recommends using a non-informative uniform prior density on standard deviation parameters when fitting hierarchical models. INLA default logGamma(1,0.00005) prior on the log precision has been also used in the current paper

obtaining similar results. However, these priors may lead to wrong results,[36] and log gamma priors have been criticized in the literature (see for example Simpson et al.[37]).
Three different model selection criteria were considered to select the best model: a modified version of the Deviance Information Criterion (DIC)[38] proposed by Plummer,[39] the Watanabe-Akaike Information Criterion (WAIC)[40] and the logarithmic score (LS).[41]

Identifiability constraints

Section 3 models are not identifiable as the intercept is also included in the P-spline smooth functions. To achieve model identifiability, the P-spline coefficients are centered. The required sum-to-zero constraints for each model when considering random walks of first order as prior distributions for the coefficients are given in Table 3.

Table 3: Sum-to-zero constraints to fit the Bayesian P-spline models.

Model	Sum-to-zero constraints
Model M1	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0;\ \sum_{a=1}^{a_m}\theta_a = 0$
Model M2	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0;\ \sum_{a=1}^{a_m}\theta_a = 0;\ \sum_{a=1}^{a_m}\theta_{sa} = 0, \forall s$
Model M3	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0,\ \sum_{a=1}^{a_m}\theta_a = 0;\ \sum_{t=1}^{t_k}\theta_{st} = 0, \forall s$
Model M4	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0;\ \sum_{a=1}^{a_m}\theta_a = 0;\ \sum_{t=1}^{t_k}\theta_{st} = 0, \forall s;\ \sum_{a=1}^{a_m}\theta_{sa} = 0, \forall s$
Model M5	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{ta} = 0$
Model M6	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{tsa} = 0, \forall s$
Model M7	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0;\ \sum_{a=1}^{a_m}\theta_a = 0;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{tsa} = 0, \forall s$
Model M8	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\theta_t = 0;\ \sum_{a=1}^{a_m}\theta_a = 0;\ \sum_{t=1}^{t_k}\theta_{st} = 0, \forall s;\ \sum_{a=1}^{a_m}\theta_{sa} = 0, \forall s;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{tsa} = 0, \forall s$
Model M9	$\sum_{s=1}^{S}\phi_s = 0;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{ta} = 0;\ \sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{tsa} = 0, \forall s$

Posterior patterns

The main and different interaction effects can be computed from the estimated rates as follows

$$
\begin{aligned}
\gamma^* &= \frac{1}{TSA}\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{a=1}^{A}\log(\hat{r}_{tsa}), \quad \phi_s^* = \frac{1}{TA}\sum_{t=1}^{T}\sum_{a=1}^{A}\log(\hat{r}_{tsa}) - \gamma^*, \\
f_t^* &= \frac{1}{SA}\sum_{s=1}^{S}\sum_{a=1}^{A}\log(\hat{r}_{tsa}) - \gamma^*, \quad f_a^* = \frac{1}{TS}\sum_{t=1}^{T}\sum_{s=1}^{S}\log(\hat{r}_{tsa}) - \gamma^*, \\
f_{sa}^* &= \frac{1}{T}\sum_{t=1}^{T}\log(\hat{r}_{tsa}) - \phi_s^* - f_a^* - \gamma^*, \quad f_{ts}^* = \frac{1}{A}\sum_{a=1}^{A}\log(\hat{r}_{tsa}) - \phi_s^* - f_t^* - \gamma^* \\
f_{ta}^* &= \frac{1}{S}\sum_{s=1}^{S}\log(\hat{r}_{tsa}) - f_t^* - f_a^* - \gamma^*, \\
f_{tsa}^* &= \log(\hat{r}_{tsa}) - \phi_s^* - f_t^* - f_a^* - f_{sa}^* - f_{ts}^* - f_{ta}^* - \gamma^*,
\end{aligned}
$$

where γ^* is the log of the overall rate in the whole country, time period and age groups, ϕ_s^*, f_t^*, f_a^* are the global spatial, temporal, and age effects, respectively; f_{sa}^*, f_{ta}^*, and f_{tsa}^* are space-age, time-age and space-time-age interactions. The terms f_{sa}^* can be interpreted as spatial patterns for age groups common to all years, the terms f_{ta}^* are temporal trends for the different age groups common to all areas, and finally, f_{tsa}^* can be seen as temporal trends for the different age groups specific for each area. This decomposition is very useful as it allows for comparing different models term by term.

5 Illustration results

P-spline models are illustrated with female breast cancer mortality in Spain from 1985 to 2010. In all models, 8 and 4 equidistant internal knots have been considered to construct the marginal cubic B-spline bases for time, $\mathbf{B}_t$, and age, $\mathbf{B}_a$, respectively. Table 4 displays the values of the different model selection criteria. To conserve space, only results for first order random walks are displayed as second order random walks did not lead to an improvement. Models based on two dimensional P-splines (M5 to M9) have been fitted using both fixed relative scale and scale invariant penalties (see Table 3). According to all model selection criteria, Model 9-SI is the best candidate. In addition, and following the suggestion of one reviewer, deviance and Pearson residuals plots (not shown here to conserve space) have been inspected. Some of the models (all models except models 6 and 9) exhibit a kind of linear tendency and large values of the residuals indicating that they may not be appropriate. Residual plots for Model 9-SI do not show any pattern and we do not observe too many large residuals. Models 9-FRS, 6-FRS, and 6-SI also lead to residuals plots without any clear pattern. However, as the model selection criteria points towards Model 9-SI, and scale invariant smooths are preferred, this is the model we finally use to analyze the data. Fitting Model M9-SI takes 45 minutes in a twin superserver with four processors Intel Xeon 6C and 96GB RAM using `R` (version 3.2.2) and the `R` package `INLA` (version 0.0-1455098891, dated 2016-02-10).

Figure 3 displays $exp(\phi_s^*)$, $exp(f_t^*)$, and $exp(f_a^*)$. 95% credible bands, made up of 95% pointwise credible intervals, are also displayed for the mean posterior temporal pattern ($exp(f_t^*)$) and mean posterior age pattern ($exp(f_a^*)$). The credible band for the age effect is very narrow. However,

Table 4: Model selection criteria for the different models. Models based on two-dimensional P-splines have been fitted using fixed relative scale and scale invariant penalties.

Model	DIC_c	$WAIC$	LS
M1	68903.1	68918.2	2.410
M2	68779.0	68832.9	2.407
M3	68711.2	68753.7	2.404
M4	68568.9	68636.5	2.400
M5-FRS (Fixed relative scale)	67819.9	67839.0	2.372
M5-SI (Scale Invariant)	67818.6	67831.5	2.372
M6-FRS (Fixed relative scale)	69927.0	69801.4	2.447
M6-SI (Scale Invariant)	67678.9	67696.1	2.367
M7-FRS (Fixed relative scale)	67511.3	67535.4	2.361
M7-SI (Scale Invariant)	67459.1	67480.9	2.359
M8-FRS (Fixed relative scale)	67686.8	67708.5	2.368
M8-SI (Scale Invariant)	67677.6	67694.7	2.367
M9-FRS (Fixed relative scale)	67313.1	67333.8	2.354
M9-SI (Scale Invariant)	**67280.2**	**67297.9**	**2.353**

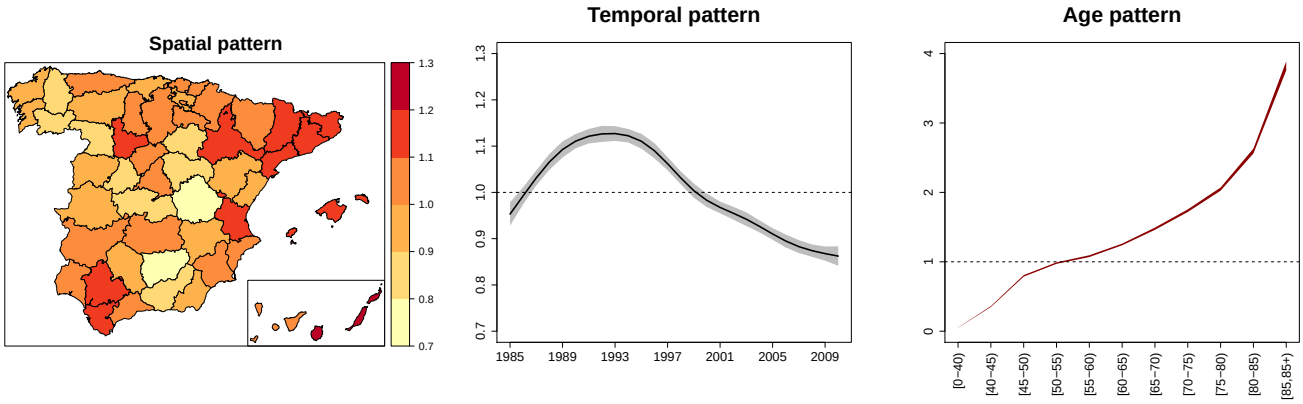

Figure 3: Female breast cancer spatial effects $exp(\phi_s^*)$ common to all years and age groups (left), temporal effects $exp(f_t^*)$ (with 95% credible bands) common to all provinces and age groups(middle), and age pattern $exp(f_a^*)$ (with 95% credible bands) common to all years and provinces (right). Values greater than one contribute to increase final rates.

this is expected as we have 1,300 observations for each age group. The spatial pattern $exp(\phi_s^*)$ is common to all years and age groups and can be interpreted as the base rate associated to a province. The temporal pattern $exp(f_t^*)$ is common to all areas and age-groups and can be

seen as a global temporal trend associated to both improvements in treatments (which generally apply to all provinces and age groups), and common health policies in Spain. The age pattern $exp(f_a^*)$ is common to all provinces and years and can describe a general evolution of mortality with age. As the small areas are defined by crossing the geographical units (provinces), the time period (years), and the age groups, these terms are not small area-specific and can be seen as a way to "borrow strength". For example, the spatial effect (ϕ_s^*) borrows information from neighboring regions, years, and age-groups.

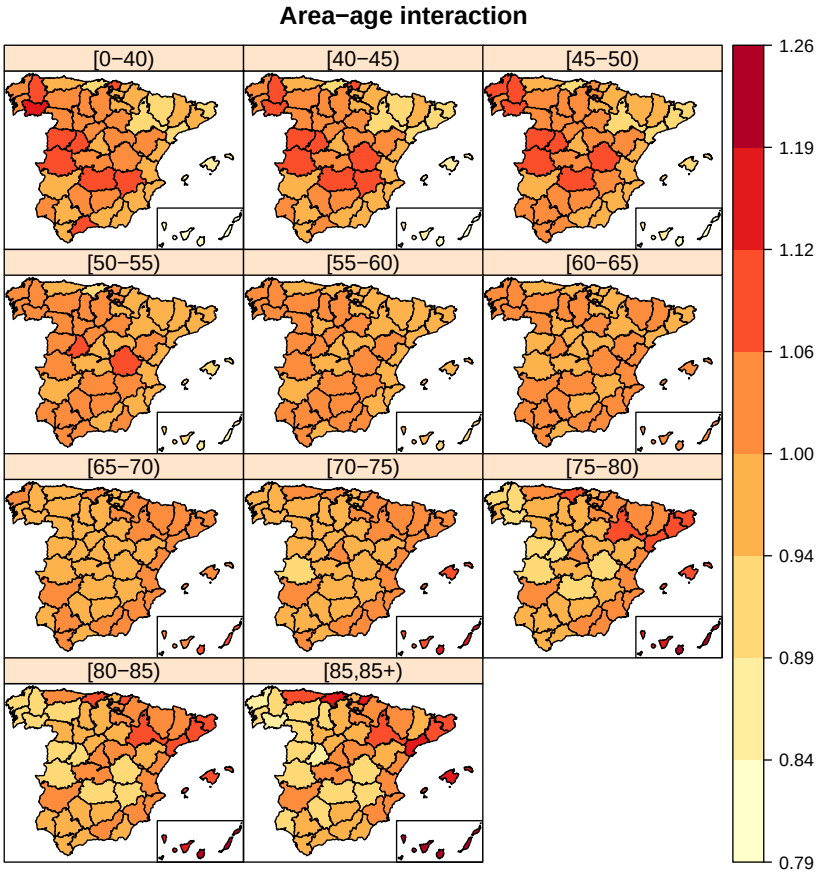

Figure 4: Area-age interaction effect in female breast cancer mortality ($exp(f_{sa}^*)$).

The area-age, age-time, and time-age interaction can be also plotted. Figure 4 displays the province-age interaction effect ($exp(f_{sa}^*)$) common to all years revealing that the age-specific contribution of each province to the total rate is different. For example, the Galician provinces (North-West) contribute to increase the final rate in the age groups $[0, 40)$, $[40, 45)$, $[45, 50)$, $[50, 55)$, $[55, 60)$, $[60, 65)$, but they contribute to decrease global rates in the oldest age groups. It is also interesting to look at the time-age interaction effect ($exp(f_{ta}^*)$) as it can be interpreted as the temporal mortality trend specific for each age group and common to all areas. Figure 5 displays the time-age interaction effect revealing decreasing trends for females aged 65 or less, and increasing trends for females aged 75 or more (credible bands are not shown for a better inspection of the trends).

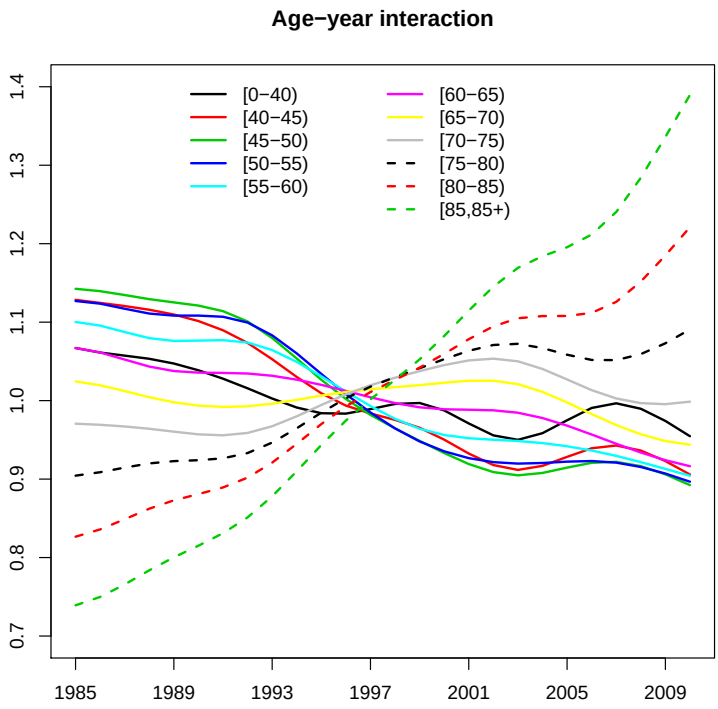

Figure 5: Time-age interaction effect in female breast cancer mortality ($exp(f^*_{ta})$).

The complete view of the temporal evolution of female breast cancer mortality rates by age group and province is displayed in Figure 6 for three provinces: Madrid, Navarra, and Huelva. In addition, 95% credible bands are also displayed. Madrid is a highly populated province in the center of the country and, as expected, credible bands for the final rates are narrower. Navarra is located in the North of Spain, and Huelva in the South. They are both low populated areas and consequently, 95% credible bands are wider. On one hand decreasing mortality trends are detected for young age groups (left column in Figure 6) although the decrease in Huelva is not as steep as in Navarra or Madrid. On the other hand, mortality trends for old age groups (right column in Figure 6) are not decreasing. In particular, the mortality trend for the oldest age group is increasing in the three provinces.
Finally, Figure 7 shows the evolution of the geographical pattern of breast cancer mortality rates for one of the youngest age group ($[45, 50)$) and one of the oldest ($[80, 85)$) to see the differences. In each figure, the cutoff value between orange and blue colors indicates the average rate of Spain over the entire study period for that particular age group. Regions colored in red (from dark to light) mean that the rates are between 1.75 and 2, 1.5 and 1.75, 1.25 and 1.5, and 1 and 1.2 times higher than the mean rate of Spain, respectively. Dark and light blue colors indicate that the rates are 1.75-2, 1.5-1.75, 1.25-1.5 and 1-1.25 times lower than the mean rate of the whole country. For the age-group [45,50), a clear reduction of mortality is observed from 1999-2000 onwards. This age-group is directly affected by the introduction of early detection programmes

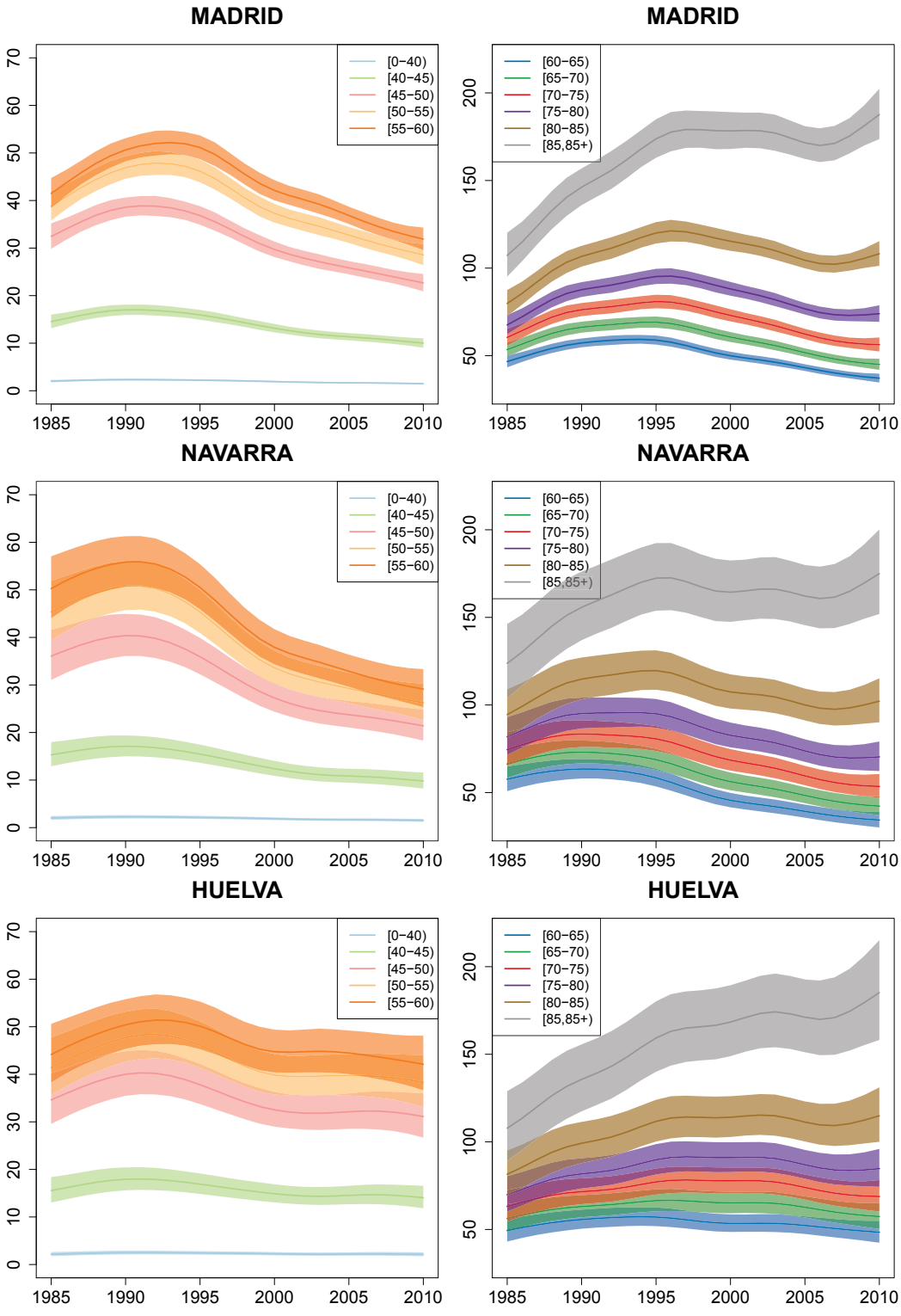

Figure 6: Temporal evolution of mortality rates and 95% credible bands by age group and province for three selected provinces: Madrid, Navarra, and Huelva. Young groups are display in the left column and older age groups are plotted in the right column.

because since 1990, women aged between 40 and 69 were selected for mammography, and an early diagnosis is generally translated into better prognosis. Nowadays these programmes have already attained 100% coverage. This map reflects that the effects of a screening programme are visible roughly in ten years-ahead. For the age group [80,85), the reduction of mortality is not as clear as in the previous group. In fact, the geographical pattern does not change that much with time, and the North-Mediterranean areas and the islands exhibit the highest rates.

6 Discussion

In the last few years, disease mapping research has evolved from spatial to spatio-temporal modelling due to the availability of high quality registers. Much of the research has been conducted within a fully Bayes approach, leading to a general acceptance of Bayesian methods in public health institutes.

It has been common practice in disease mapping to assume that age groups are similarly affected by a disease, but interactions can occur, and age-space-time models should be considered. Age-space-time specific rates can be highly variable and sophisticated statistical models are required to smooth rates and reduce variability. In this paper, the goal is to combine both P-splines for age and time, and CAR priors for space. CAR spatial random effects produce local smoothing. This approach may be very convenient in Spain where some policies related to health care and treatments are within the competence of the Autonomous Regions, and then, inequalities in health among the different provinces are likely to exist. That is the main reason why we have considered a CAR prior for the spatial random effect. In addition, Ugarte et al.,[25] show that a model with a CAR prior for the spatial effect outperforms the model with a two-dimensional P-spline to deal with spatial heterogeneity. Here, age-time interactions are modeled using two-dimensional P-splines. As far as we we know, this proposal is new in disease mapping, yet it is common in life table smoothing.[42] The final model allows the estimation of mortality temporal trends by age groups and provinces, providing valuable information to public health institutes, health researchers, and policy makers in terms of gaining knowledge about the disease, allocating funds judiciously, and evaluating screening and prevention programmes. The INLA approach has been followed for model fitting and inference speeding up computations (in comparison to McMC) and providing the full posterior distribution of the quantities of interest (the age specific space-time rates in our case). For readers interested in other fitting alternatives, Currie et al.,[43] develop an algorithm to fit tensor product models very efficiently by exploiting the grid structure of the data and provide the `R` code. In addition the `R` package `mgcv`[44] can fit models with similar structure to the ones used here, albeit with different spatial penalties.

The analysis of breast cancer mortality in Spain reveals a decline in breast cancer mortality, particularly for age groups under 70 years, although the decline is not the same in all provinces. The decrease is not very clear for the oldest age groups. This is an important point because the fitted model is flexible enough to capture different time trends according to the age groups. The differences may be due to screening programs and/or improved treatments which increase survival and reduce mortality in young age groups. The results are consistent with other studies analysing the role of the screening programmes and improvements in treatment in European countries such as France, Italy, Norway, Slovakia, Slovenia, Switzerland, Czech Republic, Denmark, or Estonia.[21] In these countries a significant decline in breast cancer mortality was found in the majority of the age-groups except in women older than 65 years. An analysis of age-

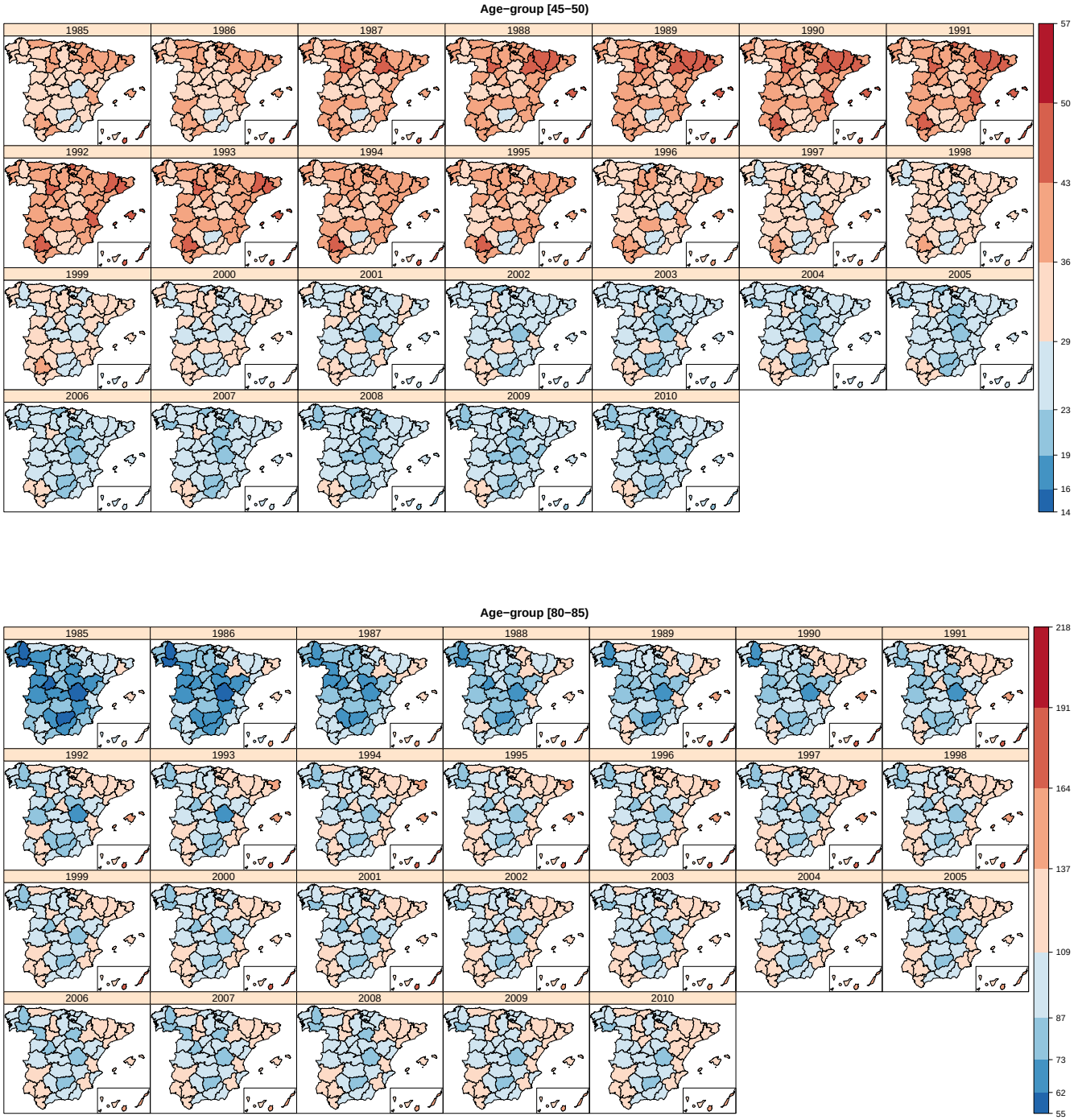

Figure 7: Age-specific breast cancer mortality patterns in Spain (1985-2010) for the age groups $[45-50)$ (top), and $[80-85)$‘(bottom).

related breast cancer mortality rates for the United States and England-Wales[45] reports similar results concluding that mortality trends for middle-aged women have been declining since the 1990s, whereas mortality rates for women older than 80 are expected to increase or stabilize.

Our results agree with Holford et al.[46] and Clèries et al.,[47] where the decrease in mortality rates in older women is not observed or it is observed later than in youngest age-groups regardless of screening programes or advances in treatment. This suggests less efficient treatment schemes in the oldest age groups[48].
To sum up, the models presented in this work are very flexible and convenient when different temporal mortality trends by age groups and regions are likely to occur. The models can be implemented in `R-INLA`, but caution is recommended to deal with identifiability constraints. It is also important to highlight that when variables are measured in different scales, scale invariant penalties leading to scale invariant smooths are preferred[28]. The methodology is very useful to gain a general perspective about the evolution of mortality, and therefore a better evaluation of prevention programmes implemented in each region.

Acknowledgements

This work has been supported by the Spanish Ministry of Economy and Competitiveness (project MTM2014-51992-R), and by the Health Department of the Navarre Government (Project 113, Res.2186/2014). We would like to thank the National Epidemiology Center (area of Environmental Epidemiology and Cancer) for providing the data, originally created by the Spanish Statistical Office.

References

1. Ugarte MD, Ibáñez B, and Militino AF. Modelling risks in disease mapping. *Stat Methods Med Res* 2006; **15**: 21-35.

2. Eilers PHC and Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci* 1996; **11**: 89-121.

3. Lang S and Brezger A. Bayesian P-Splines. *J Comput Graph Stat* 2004; **13(1)**: 183-212.

4. MacNab YC and Dean CB. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics* 2001; **57(3)**: 949-956.

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

6. Ugarte MD, Goicoa T, Etxeberria J, and Militino AF. A P-spline ANOVA-type model in space-time disease mapping. *Stoch Env Res Risk A* 2012; **26(6)**: 835-845.

7. Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; **88(421)**: 9-25.

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

9. MacNab YC. Spline smoothing in Bayesian disease mapping. *Environmetrics* 2007; **18(7)**: 727-744.

10. Bauer C, Wakefield J, Rue H, Self S, Feng A, and Wang Y. Bayesian penalized spline models for the analysis of spatio-temporal count data. *Stat Med* 2016; **35(11)**: 1848-1865.

11. Schrödle B and Held L. A primer on disease mapping and ecological regression using INLA. *Computation Stat* 2011; **26(2)**: 241-258.

12. Rue H, Martino S, and Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J Roy Stat Soc B* 2009; **71(2)**: 319-392.

13. Sun D, Tsutakawa RK, Kim H, and He Z. Spatio-temporal interaction with disease mapping. *Stat Med* 2000; **19(15)**: 2015-2035.

14. Zhang S, Sun D, He CZ, and Schootman M. A Bayesian semi-parametric model for colorectal cancer incidences. *Stat Med* 2006; **25(2)**: 285-309.

15. Papoila AL, Riebler A, Amaral-Turkman A, São-João R, Ribeiro C, Geraldes C, and Miranda A. Stomach cancer incidence in Southern Portugal 1998-2006: a spatio-temporal analysis. *Biometrical Journal* 2014, **56(3)**: 403-415.

16. Etxeberría J, Goicoa T, López-Abente G, Riebler A, and Ugarte MD . Spatial gender-age-period-cohort analysis of pancreatic cancer mortality in Spain (1990-2013). *Plos one* 201); **12**(2): 1-12.

17. Goicoa T, Ugarte MD, Etxeberría J, and Militino AF. Age-space-time CAR models in Bayesian disease mapping. *Stat Med* 2016; **35(14)**: 2391-2405.

18. International Organization Health. *ICD-10a. International Statistical Classification of Diseases and related Health problems.* `http://www.who.int/classifications/icd/en/`

19. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013; **49(6)**: 1374-1403.

20. Pollán M, García-Mendizábal MJ, Gómez BP, Aragonés N, Lope V, Pastor R, et al. Situación epidemiológica del cáncer de mama en España. *Psicooncología* 2007; **4(2)**: 231-248.

21. Botha JL, Bray F, Sankila R, Parkin DM. Breast cancer incidence and mortality trends in 16 European countries. *Eur J Cancer* 2003; **39(12)**: 1718-1729.

22. Ascunce N, Salas D, Zubizarreta R, Almazán R, Ibáñez J, Ederra M. Cancer screening in Spain. *Ann Oncol* 2010; **21(3)**: 43-51.

23. Álvaro-Meca A, Debón A, Gil-Prieto R, Gil de Miguel S. Breast cancer mortality in Spain: has it really declined for all age groups? *Public Health* 2012; **126(10)**: 891-895.

24. Ugarte MD, Goicoa T, Etxeberria J, Militino AF, and Pollán M. Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975-2005). *Ann Epidemiol* 2010; **20(12)**, 906-9016.

25. Ugarte MD, Adin A, and Goicoa T. One-dimensional, two-dimensional, and three dimensional B-splines to specify space-time interactions in Bayesian disease mapping: Model fitting and model identifiability. *Spatial Statistics* 2017, in press. http://dx.doi.org/10.1016/j.spasta.2017.04.002

26. Leroux BG, Lei X, and Breslow N. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials* (eds Halloran M and Berry D) 1999; 179-192. New York: Springer-Verlag.

27. Rue H and Held L. *Gaussian Markov random fields: theory and applications.* Chapman & Hall/CRC, Boca Raton, 2005.

28. Wood SN, Scheipl F, and Faraway JJ. Straightforward intermediate rank tensor product smoothing in mixed models. *Stat Comput* 2013; **23(3)**: 341-360.

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

30. Schrödle B. and Held L. Spatio-temporal disease mapping using INLA. *Environmetrics* 2011; **22(6)**: 725-734.

31. Ugarte MD, Adin A, Goicoa T, and Militino AF. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat Methods Med Res* 2014; **23(6)**:507-530.

32. R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/` (2015).

33. Martino S and Rue H. *Implementing approximate Bayesian inference using integrated nested Laplace approximation: A manual for the inla program.* Department of Mathematical Sciences, NTNU, Norway, `http://www.math.ntnu.no/~hrue/GMRFLib/manual.pdf` (2010).

34. Goicoa T, Adin A, Ugarte MD, and Hodges J. In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment* 2017, in press DOI 10.1007/s00477-017-1405-0.

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

36. Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, and Watjou K. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology* 2015; **14-15**: 45-54.

37. Simpson D, Rue H, Riebler A, Martins TG, and Sorbye SH. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Stat Sci* 2017; **32**: 1-28. doi:10.1214/16-STS576.

38. Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A. Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 2002, **64(4)**: 583-639.

39. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics* 2008; **9(3)**:523-539.

40. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 2010; **11**: 3571-3594.

41. Gneiting T and Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* 2007; **102(477)**:359-378.

42. Currie ID, Durbán M, and Eilers PH. Smoothing and forecasting mortality rates. *Stat Model* 2004; **4**: 279-298.

43. Currie ID, Durbán M, and Eilers PH. Generalized linear array models with applications to multidimensional smoothing. *J Roy Stat Soc B* 2006; **68**: 1-22.

44. Wood SN. Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation 2017. R package version 1.8-17.

45. Erbas B, Akram M, Gertig DM, English D, Hopper JL, Kavanagh AM, and Hyndman R. Using Functional Data Analysis Models to Estimate Future Time Trends in Age-Specific Breast Cancer Mortality for the United States and England-Wales. *J Epidemiol* 2010; **20(2)**: 159-165.

46. Holford TR, Cronin KA, Mariotto AB, and Feuer EJ. Changing patterns in breast cancer incidence trends. *J Natl Cancer Inst Monogr* 2006; **36**: 19-25.

47. Clèries R, Ribes J, Esteban L, Martínez JM, and Borrás JM. Time trends of breast cancer mortality in Spain during the period 1977-2001 and Bayesian approach for projections during 2002-2016. *Ann Oncol* 2006; **17(12)**: 1783-1791.

48. Renard F, Van Eycken L, and Arbyn M. High burden of breast cancer in Belgium: recent trends in incidence (1999-2006) and historical trends in mortality (1954-2006). *Arch Public Health* 2011; **69**: 2.

Appendix: R-INLA code

The R code to fit the scale invariant model M9 described in Section 3 with INLA is detailed below. First, the data with the number of deaths (`observed`), the population at risk (`population`), time period, geographical area, and age-group are uploaded. Note that data must be ordered according to the Kronecker product defined for models M1-M9.

```
> library(INLA)
> library(splines)
>
> Data <- read.table("...")
>
> T <- "number of time periods"
> S <- "number of geographical areas"
> A <- "number of age-groups"
```

Then, the marginal B-spline bases for time and age ($\mathbf{B}_t$ and $\mathbf{B}_a$ respectively) are constructed. When analyzing breast cancer mortality data in Section 5, cubic B-spline bases with 8 equidistant internal knots for the time covariate and 4 equidistant internal knots for the age covariate have been considered.

```
> p <- 3       ## Cubic B-splines
> q.time <- 7  ## Number of internal intervals for time (knots-1)
> q.age <- 3   ## Number of internal intervals for age (knots-1)
>
> ## Marginal basis for time-periods ##
> xt <- "time covariate"
> xt <- (xt-min(xt))/(max(xt)-min(xt))  ## Scale time to lie in [0,1]
>
> dis.t <- (max(xt)-min(xt))/q.time
> xl.t <- min(xt)-dis.t*0.05
> xr.t <- max(xt)+dis.t*0.05
> dx.t <- (xr.t-xl.t)/q.time
> knots.t <- seq(xl.t-p*dx.t, xr.t+p*dx.t, by=dx.t)
>
> Bt <- spline.des(knots.t,xt,p+1)$design  ## B-spline basis for time
> kt <- ncol(Bt)
>
> ## Marginal basis for age-groups ##
> xa <- "age covariate"
> xa <- (xa-min(xa))/(max(xa)-min(xa))  ## Scale age to lie in [0,1]
>
> dis.a <- (max(xa)-min(xa))/q.age
> xl.a <- min(xa)-dis.a*0.05
> xr.a <- max(xa)+dis.a*0.05
> dx.a <- (xr.a-xl.a)/q.age
> knots.a <- seq(xl.a-p*dx.a, xr.a+p*dx.a, by=dx.a)
>
> Ba <- spline.des(knots.a,xa,p+1)$design  ## B-spline basis for age
> ka <- ncol(Ba)
```

The structure matrices for the main random effects must be also defined. The LCAR prior for the spatial random effect $\boldsymbol{\phi}_s = (\phi_1, \ldots, \phi_S)'$ can be implemented in INLA using the `"generic1"` model.[31]

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

where "spatial_nb.inla" is the inla.graph object containing the neighboring structure of the geographical areas ($S = 50$ provinces in the illustration). First order random walk priors have been considered for the coefficients of the B-spline bases for time $\boldsymbol{\theta}_t = (\theta_1, \ldots, \theta_{t_k})'$ and age $\boldsymbol{\theta}_a = (\theta_a, \ldots, \theta_{a_m})'$ covariates.

```
> Dt <- diff(diag(kt),differences=1)
> Pt <- t(Dt)%*%Dt
>
> Da <- diff(diag(ka),differences=1)
> Pa <- t(Da)%*%Da
```

Next, the hyperprior distributions described in Section 4 are implemented.

```
> sdunif="expression:
>   logdens=-log_precision/2;
>   return(logdens)"
>
> lunif = "expression:
>   a = 1;
>   b = 1;
>   beta = exp(theta)/(1+exp(theta));
>   logdens = lgamma(a+b)-lgamma(a)-lgamma(b)+
>             (a-1)*log(beta)+(b-1)*log(1-beta);
>   log_jacobian = log(beta*(1-beta));
>   return(logdens+log_jacobian)"
```

Model M9 can be expressed in matrix form as

$$\log \mathbf{r} \simeq (\mathbf{1}_T \otimes \mathbf{1}_S \otimes \mathbf{1}_A)\gamma + (\mathbf{1}_T \otimes \mathbf{I}_S \otimes \mathbf{1}_A)\boldsymbol{\phi}_s + (\mathbf{B}_t \otimes \mathbf{1}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{ta} + (\mathbf{B}_t \otimes \mathbf{I}_S \otimes \mathbf{B}_a)\boldsymbol{\theta}_{tsa},$$

and the design matrices of the random effects are defined.

```
> ones.T <- matrix(1,T,1)
> ones.S <- matrix(1,S,1)
> ones.A <- matrix(1,A,1)
>
> Xs <- kronecker(ones.T,kronecker(diag(S),ones.A))
> Xta <- kronecker(Bt,kronecker(ones.S,Ba))
> Xtsa <- kronecker(Bt,kronecker(diag(S),Ba))
```

Finally, the formula object is defined to be used with a call to the inla() function

```
## List containing the variables in the model ##
> Data.inla <- list(O=observed, E=population,
>             Intercept=c(1,rep(NA,S+kt*ka+kt*S*ka)),
>             ID.area=c(NA,1:S,rep(NA,kt*ka+kt*S*ka)),
>             ID.year.age=c(rep(NA,1+S),1:(kt*ka),rep(NA,kt*S*ka)),
>             ID.year.area.age=c(rep(NA,1+S+kt*ka),1:(kt*S*ka)))
```

```
>
> ## Penalty matrices ##
> R1 <- list(inla.as.sparse(kronecker(Pt,diag(ka))),
>            inla.as.sparse(kronecker(diag(kt),Pa)))
>
> R2 <- list(inla.as.sparse(kronecker(Pt,kronecker(diag(S),diag(ka)))),
>            inla.as.sparse(kronecker(diag(kt),kronecker(diag(S),Pa))))
>
> ## Sum-to-zero constraints ##
> A.constr <- kronecker(matrix(1,1,kt),kronecker(diag(S),matrix(1,1,ka)))
>
> ## Model formula ##
> M9.SI<- O ~ -1 + Intercept +
>       f(ID.area, model="generic1", Cmatrix=Q.Leroux, constr=TRUE,
>         hyper=list(prec=list(prior=sdunif),beta=list(prior=lunif))) +
>       f(ID.year.age, model="generic3", Cmatrix=R1, constr=TRUE,
>         hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)))+
>       f(ID.year.area.age, model="generic3", Cmatrix=R2, constr=TRUE,
>         hyper=list(prec1=list(prior=sdunif),prec2=list(prior=sdunif)),
>         extraconstr=list(A=A.constr, e=rep(0,S)))
>
> ## Call to the inla() function ##
> inla(M9.SI, family="poisson", data=Data.inla, E=E,
>   control.predictor=list(compute=T, A=cBind(rep(1,T*S*A),Xs,Xta,Xtsa)),
>   control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
>   control.inla=list(strategy="simplified.laplace"))
```

where `R1` and `R2` define the scale invariant penalty matrices for the time-age two-dimensional P-spline coefficients $\boldsymbol{\theta}_{ta}$ and $\boldsymbol{\theta}_{tsa}$ respectively (see the expressions of $\mathbf{P}_{ta}$ and $\mathbf{P}_{tsa}$ in Table 2). The sum-to-zero constraints that make the model identifiable (see Table 3) are specified through the `constr=TRUE` and `extraconstr` arguments. In model M9, the constraints over the coefficients $\boldsymbol{\theta}_{tsa}$ are expressed as

$$\sum_{t=1}^{t_k}\sum_{a=1}^{a_m}\theta_{tsa}=0, \quad \text{for} \quad s=1,\ldots,S \iff (\mathbf{1}_{t_k}^{'}\otimes\mathbf{I}_S\otimes\mathbf{1}_{a_m}^{'})\boldsymbol{\theta}_{tsa}=\mathbf{0},$$

and are passed through the `extraconstr` argument.