

Age–space–time CAR models in Bayesian disease mapping

T. Goicoa,^{a,b,c} M. D. Ugarte,^{a,b} J. Etxeberria^{a,b,d} and A. F. Militino^{a,b*†}

Mortality counts are usually aggregated over age groups assuming similar effects of both time and region, yet the spatio-temporal evolution of cancer mortality rates may depend on changing age structures. In this paper, mortality rates are analyzed by region, time period and age group, and models including space–time, space–age, and age–time interactions are considered. The integrated nested Laplace approximation method, known as INLA, is adopted for model fitting and inference in order to reduce computing time in comparison with Markov chain Monte Carlo (McMC) methods. The methodology provides full posterior distributions of the quantities of interest while avoiding complex simulation techniques. The proposed models are used to analyze prostate cancer mortality data in 50 Spanish provinces over the period 1986–2010. The results reveal a decline in mortality since the late 1990s, particularly in the age group [65, 70), probably because of the inclusion of the PSA (prostate-specific antigen) test and better treatment of early-stage disease. The decline is not clearly observed in the oldest age groups. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: INLA; interaction models; mortality rates

1. Introduction

Aggregating counts over all age groups inside a region in a specific period of time is a common practice to study age-standardized mortality (or incidence) risks or rates. This means that region and temporal effects are supposed to be the same over the age groups, and hence, a single estimate is provided. This is a simple framework of analysis, and it should be regarded as a starting point [1]. If a single estimated rate is provided for each area and time, it is implicitly assumed that age groups are similarly affected by the disease, and this may not be necessarily true as the effects on children or elderly people could be more pronounced than on youngsters. Therefore, if the age groups are not equally affected by the disease, this practice could lead to misleading conclusions. Conditional autoregressive (CAR) models have been and still are one of the most popular approaches to disease mapping since the pioneering research by Besag et al. [2]. The more common assumption is based on the same spatial and spatio-temporal effect on the age groups [3–7]. It can be argued that in certain diseases there is no biological reason to assume a different age effect for each area [8], but in other cases region effects such as pollution, may have different consequences for distinct age groups [9]. Models incorporating age effects usually follow the proportional assumption [10], whereby separate area and age effects multiply to produce area–age rates. In any case, regardless of whether area–age interactions are considered, the necessity of dealing with the potential age effect in these models is clear, because when the evolution of the disease is not the same among the age groups, age-specific rates within each region should be provided. Embedding the effect of age in mortality rates is not new in spatial or spatio-temporal disease mapping, yet it has been only partially studied. For example, Nandram et al. [11] consider local health areas within larger regions to study mortality rates

^aDepartment of Statistics and O. R. Universidad Pública de Navarra, Campus de Arrosadia, 31006 Pamplona, Spain

^bInstitute for Advanced Materials (INAMAT), Universidad Pública de Navarra, Campus de Arrosadia, 31006 Pamplona, Spain

^cResearch Network on Health Services in Chronic Diseases (REDISSEC), Spain

^dConsortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Spain

*Correspondence to: A. F. Militino, Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain.

†E-mail: militino@unavarra.es

for chronic obstructive pulmonary disease. They include interactions between the larger regions, but the age effects are modeled as independent random effects. Dean et al. [9] introduce a spatial effect with a CAR prior and an unstructured random effect to deal with area–age interactions. They also derive a score test to assess if the interaction term is statistically significant, something extremely useful as this test indicates whether the simplest model is adequate or not. Silva and Dean [1] use B-splines to model the effect of age on lung cancer incidence rates where both additive and area–age interaction models are considered. MacNab and Dean [12] study the temporal evolution of risks in British Columbia using common and area-specific temporal B-splines. They also comment on the possibility of using B-splines to model the age effect, but only additively. Age–space–time interactions have scarcely been analyzed in specific cases. Sun et al. [13] propose an age–space–time model including the age as a fixed effect. The temporal trend is linear and specific for each area and age group. However, they do not consider age–space interactions, and a linear trend may be very restrictive in practice. Zhang et al. [14] propose a Gaussian demographic spatio-temporal CAR process. Their model is quite flexible and gender–age–space–time interactions are modeled through the covariance matrix. However, the interpretation is more difficult as the gender, age, temporal, and spatial components are not explicitly defined.

In this paper, mortality rates by region, time, and age groups are provided to gain knowledge about the spatial variation and temporal evolution of mortality. The main goal of this work is to detect spatio-temporal patterns according to the different age groups. These smooth rates allow us to compare mortality figures among regions and age groups in different periods. In a first approach, we consider additive models with CAR prior for space, and first order random walks for both time and age effects. However, these models may be very restrictive in practice as interactions usually occur. Therefore, we later choose interaction models of Type IV described by Knorr-Held [3], this type of interaction being more suitable if temporal trends are different for far apart regions but tend to be similar for adjacent regions. Estimation will be carried out using Bayesian methodology. Specifically, integrated nested Laplace approximations (INLA) will be adopted for approximate Bayesian inference [15].

The rest of the paper is organized as follows. Section 2 describes the proposed CAR models. Section 3 gives a brief summary of the INLA methodology. Section 4 shows the results for the case study of prostate cancer in Spain over the period 1986–2010. Finally, Section 5 closes the paper with a discussion.

2. Age–space–time conditional autoregressive models

In this section, different CAR models are proposed to study the temporal evolution of geographical patterns of mortality rates by age group. They range from the simplest additive models to more complex alternatives including interactions. In the following, the spatial domain corresponds to the whole country and it is divided into I small areas or provinces, indexed by $i = 1, \dots, I$. The age groups of interest are denoted by $j = 1, \dots, J$. Suppose that for each small area i and each age group j , data are available in different years $t = 1, \dots, T$. Let Y_{ijt} , n_{ijt} , and r_{ijt} be the number of deaths of a particular disease, the population at risk, and the mortality rate, respectively, in region i , age group j , and time t . Then, conditional on the rate, the number of deaths is assumed to follow a Poisson distribution

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

where the term $\log n_{ijt}$ is an offset. Depending on how $\log r_{ijt}$ is specified, different models arise. Let us first start with a simple additive model

$$\text{Model 1: } \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t, \quad (1)$$

where β is an overall rate, and ϕ_i , δ_j , and γ_t , are spatial, age, and temporal random effects, respectively. Denoting by $\boldsymbol{\phi} = (\phi_1, \dots, \phi_I)'$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_J)'$, and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_T)'$ the vectors of spatial, age, and temporal effects, they are assumed to follow multivariate normal distributions, that is,

$$\begin{aligned} \boldsymbol{\phi} &\sim N(\mathbf{0}, \sigma_\phi^2 \mathbf{D}_\phi), \quad \mathbf{D}_\phi = (\lambda_\phi \mathbf{Q}_\phi + (1 - \lambda_\phi) \mathbf{I}_\phi)^{-1}, \\ \boldsymbol{\delta} &\sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{D}_\delta), \quad \mathbf{D}_\delta = \mathbf{Q}_\delta^-, \\ \boldsymbol{\gamma} &\sim N(\mathbf{0}, \sigma_\gamma^2 \mathbf{D}_\gamma), \quad \mathbf{D}_\gamma = \mathbf{Q}_\gamma^-. \end{aligned} \quad (2)$$

Here \mathbf{Q}_ϕ is the spatial neighborhood matrix defined by adjacency, where two areas are considered neighbors if they share a common border. \mathbf{I}_ϕ is an identity matrix, and λ_ϕ is a spatial smoothing parameter.

When λ_ϕ is equal to 1, all the variability is spatially structured. If it is equal to 0, all the variability is spatially unstructured. This model has been proposed by Leroux et al. [16], and it presents advantages over the commonly used convolution model by Besag et al. [2]. One of the most important advantages is that it does not lead to negative correlations for regions located further apart [17]. \mathbf{Q}_γ and \mathbf{Q}_δ are structure matrices corresponding to first order random walks for time and age, respectively [18, p. 95], and the symbol $-$ denotes the Moore–Penrose generalized inverse. Note that a first order random walk is a CAR model with the neighborhood matrix given by the structure matrix. Next, models incorporating space–age, space–time, and age–time interactions are considered. Namely,

$$\begin{aligned} \text{Model 2: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{ij}^1, \\ \text{Model 3: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{it}^2, \\ \text{Model 4: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{jt}^3, \\ \text{Model 5: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{ij}^1 + \zeta_{it}^2, \\ \text{Model 6: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{ij}^1 + \zeta_{jt}^3, \\ \text{Model 7: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{it}^2 + \zeta_{jt}^3, \\ \text{Model 8: } & \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{ij}^1 + \zeta_{it}^2 + \zeta_{jt}^3. \end{aligned}$$

In these models, ζ_{ij}^1 , ζ_{it}^2 , and ζ_{jt}^3 are random effects dealing with space–age, space–time, and age–time interactions. The vectors $\zeta^1 = (\zeta_1^1, \dots, \zeta_{IJ}^1)'$, $\zeta^2 = (\zeta_1^2, \dots, \zeta_{IT}^2)'$, and $\zeta^3 = (\zeta_1^3, \dots, \zeta_{JT}^3)'$ are assumed to follow multivariate normal distributions with covariance matrices given by the Kronecker product of the marginal covariances such that

$$\begin{aligned} \zeta^1 & \sim N(\mathbf{0}, \sigma_{\phi\delta}^2 \mathbf{D}_{\phi\delta}), \quad \mathbf{D}_{\phi\delta} = \mathbf{Q}_\phi^- \otimes \mathbf{Q}_\delta^-, \\ \zeta^2 & \sim N(\mathbf{0}, \sigma_{\gamma\phi}^2 \mathbf{D}_{\gamma\phi}), \quad \mathbf{D}_{\gamma\phi} = \mathbf{Q}_\gamma^- \otimes \mathbf{Q}_\phi^-, \\ \zeta^3 & \sim N(\mathbf{0}, \sigma_{\gamma\delta}^2 \mathbf{D}_{\gamma\delta}), \quad \mathbf{D}_{\gamma\delta} = \mathbf{Q}_\gamma^- \otimes \mathbf{Q}_\delta^-. \end{aligned}$$

The interaction terms are completely structured, and they correspond to Type IV interactions as described by Knorr-Held [3]. That is, the evolution of rates with age and temporal trends tend to be similar in neighboring regions, and the same happens with temporal trends in contiguous age groups. A final model including a triple age–space–time interaction is considered

$$\text{Model 9: } \log r_{ijt} = \beta + \phi_i + \delta_j + \gamma_t + \zeta_{ij}^1 + \zeta_{it}^2 + \zeta_{jt}^3 + \zeta_{ijt}^4,$$

where the term ζ_{ijt}^4 is the random effect accounting for the triple age–space–time interaction. Then, the vector $\zeta^4 = (\zeta_1^4, \dots, \zeta_{JIT}^4)'$ is assumed to follow a multivariate normal distribution with covariance matrix given by the Kronecker product of the marginal covariances, that is,

$$\zeta^4 \sim N(\mathbf{0}, \sigma_{\gamma\phi\delta}^2 \mathbf{D}_{\gamma\phi\delta}), \quad \mathbf{D}_{\gamma\phi\delta} = \mathbf{Q}_\gamma^- \otimes \mathbf{Q}_\phi^- \otimes \mathbf{Q}_\delta^-.$$

Multiplicative interactions with covariance matrices defined as the Kronecker product of marginal covariance matrices have been suggested by Clayton [19]. The rationale for this proposal is that if the main effects have identity covariance matrices, so has the interaction term, and if the priors for the main effects are random walk-priors, then the structure matrix for the interaction term can be represented by an undirected conditional-independence graph. These interactions can cope with different situations. Here, for simplicity, only Type IV interactions are considered, but other interactions proposed by Knorr-Held [3] can be easily included, as it has been done by Ugarte et al. [20] in a spatio-temporal setting. For example, let us consider the spatio-temporal interaction term ξ_{it} . A Type I interaction assumes that all ξ_{it} 's are independent, that is, they do not have any structure in space and time. Type II interactions are appropriate if temporal trends do not have any spatial structure. Type III interactions produce different spatial patterns for each year without any temporal structure, and finally, in Type IV interactions, ξ_{it} 's are structured in space and time and are suitable if temporal trends from neighboring regions are likely to be similar. So they comprise a wide range of situations, from totally independent effects to completely structured random effects in space and time. The interactions are similarly defined for space–age and time–age terms.

For example, a Type IV interaction for space–age and time–age means that age patterns are similar in neighboring regions, and temporal trends are likely to be similar for contiguous age groups, respectively.

3. Model fitting

Most of the research in disease mapping involves generalized linear mixed models (GLMM) requiring approximate methods for model fitting and inference. Penalized quasi-likelihood [21] have been widely used within a frequentist or empirical Bayes approach to provide point estimates. With the advent of modern computers, a full Bayes approach using Markov chain Monte Carlo (McMC) methods has become more and more popular as the whole posterior distribution of the quantity of interest can be obtained. However, when models are complex, McMC techniques are not easy to use. In particular, algorithms have to be carefully chosen [22, 23], they can lead to a large Monte Carlo error, and computational time may be long if samples are highly correlated [24].

In recent years, a new methodology relying on integrated nested Laplace approximations has been proposed [15] for latent Gaussian fields. It is intended to overcome some of the problems of the McMC techniques, mainly when the Gaussian fields are Gaussian Markov random fields with sparse precision matrices, and the number of hyperparameters is not large. The goal is to estimate the marginal posterior distribution of the elements in the Gaussian field by approximating the marginal posterior distribution of the hyperparameters and the posterior marginal distribution of the elements in the latent Gaussian field, given the hyperparameters. To estimate this latter distribution, Rue et al. [15] propose three alternatives: a Gaussian approximation, a full Laplace approximation, and a simplified Laplace approximation. The first one is the simplest and fastest, but there can be errors in the location or because of the lack of skewness [25]. The second one is very precise, but it can be very time consuming (by time consuming, we refer to computing time), so that the third approach is used along the paper because it can be very accurate in less computational time. By very accurate, we mean that posterior distributions are practically equal to those obtained with McMC methods. In general, INLA is much faster than McMC methods (see Schrödle et al. [24] for a comparison of INLA and McMC).

Models 1 to 9 are fitted using INLA. INLA has been implemented in a package written in C [26], with a user interface in the free software R [27], called R-INLA. This can be downloaded from the web page <http://www.r-inla.org/> where information, examples, and a user's forum are available. Some interesting spatial and spatio-temporal applications with INLA can be found in Blangiardo et al. [28]. So far, R-INLA does not have a direct option to fit the model for the spatial random effects proposed by Leroux et al. [16], but it can be easily implemented using the generic1 model with the following precision matrix

$$\mathbf{R} = \tau \left(\mathbf{I}_n - \frac{\beta_0}{\lambda_{max}} \mathbf{C} \right),$$

where λ_{max} is the maximum eigenvalue of \mathbf{C} , which allows β_0 to be in the range (0, 1). In our case, \mathbf{C} is defined as

$$\mathbf{C} = \mathbf{I}_n - \mathbf{Q}_\phi = \begin{cases} -m_i + 1, & i = j \\ 1, & i \sim j \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where m_i is the number of neighbors of the i th area, and λ_{max} is equal to 1 (see [20] for more details). To include the temporal and age random effects, R-INLA has implemented specific models for random walks. Random walks of order 1 (model="rw1") have been chosen for simplicity. An important key point is that sum to zero constraints have to be used to guarantee identifiability of the interaction terms ζ^k , $k = 1, 2, 3, 4$. If \mathbf{Q}_{ζ^k} denotes the precision matrix of the interaction term ζ^k , \mathbf{Q}_{ζ^k} is not of full rank, and linear constraints $\mathbf{A}_{\zeta^k} \zeta^k = \mathbf{0}$ are needed. Here, the matrix \mathbf{A}_{ζ^k} consists of those eigenvectors of \mathbf{Q}_{ζ^k} , which span the null space (for more details, see [29]), and the number of linear constraints is equal to the rank deficiency of the precision matrix.

Another important issue in Bayesian inference is the hyperprior distributions, as they can affect the posterior distribution. In our setting, hyperparameters are the spatial smoothing parameter λ_ϕ , and the precision parameters $\tau_\phi = 1/\sigma_\phi^2$, $\tau_\gamma = 1/\sigma_\gamma^2$, $\tau_\delta = 1/\sigma_\delta^2$, $\tau_{\phi\delta} = 1/\sigma_{\phi\delta}^2$, $\tau_{\phi\gamma} = 1/\sigma_{\phi\gamma}^2$, $\tau_{\gamma\delta} = 1/\sigma_{\gamma\delta}^2$, and

$\tau_{\phi\gamma\delta} = 1/\sigma_{\phi\gamma\delta}^2$. The hyperprior distributions for the spatial components are $\log \tau_\phi \sim \text{logGamma}(1, 0.01)$ and $\text{logit}(\lambda_\phi) \sim \text{logitbeta}(1, 1)$. A non-informative prior is considered for λ_ϕ as we do not have information about the strength of the spatial dependence. For the rest of precision parameters, minimally informative priors $\log \tau_{..} \sim \text{logGamma}(1, 0.00005)$ are considered. In addition, instead of logGamma distributions, non-informative uniform priors on the positive real line have been considered for all the standard deviations obtaining practically identical final rate estimates. For the fixed effect β , a Gaussian prior with mean 0 and variance 1000 is used. The priors chosen in this paper are similar to those used in Ugarte et al. [20], where an extensive sensitivity analysis was carried out in a spatio-temporal setting. Finally, to select the best model among the different proposals, the DIC information criterion [30], a modified version of the DIC [31], and the logarithmic score [32] are considered.

The full code to fit Model 8, the one that will be selected in the case study, is provided in the Appendix.

4. Case study

In this section, an in depth analysis of prostate cancer mortality in Spain is conducted. Data are available in 50 provinces (small areas) and 9 age groups during 25 years between 1986 and 2010. As prostate cancer mainly affects aged people, age groups are defined as [-50), [50, 55), [55, 60), [60, 65), [65, 70), [70, 75), [75, 80), [80, 85), and [+85). Prostate cancer is the second leading cancer in the world and it is ranked in the sixth place in cancer mortality [33]. It is the third most common cause of cancer mortality in Spain after lung and colorectal cancer with an average of about 5000 deaths per year. During the target period, a total of 127,910 deaths were recorded, and the overall crude rate by 100,000 male inhabitants in the whole period is 25.50. A preliminary analysis of crude rates by year in the whole country reveals an increase in rates from the beginning of the period until 1998, when they start to decrease. Figure 1 displays the evolution of crude rates according to the age groups (left) and according to the years (right) for the 50 Spanish provinces. The relevant time axis in the right picture is period rather than cohort, because previous research about prostate cancer shows a cohort effect on incidence, but not on mortality [34]. Two provinces have been colored in both figures to highlight the space-age and the space-time interactions. Álava (north of Spain) and Madrid (center of Spain) are displayed in red and green, respectively, on the left picture, whereas Jaén (south of Spain) and Madrid are displayed in red and green on the right picture. The crossing lines in both figures indicate age-time and space-time interactions.

Table I displays the total number of deaths and the crude rates by 100,000 male inhabitants by age group. These measures present a different behavior of mortality by age group. In particular, rates increase with age, and this upward trend is steeper in older age groups. This pattern is common when analyzing mortality, but parabolic shapes can be found if incidence is considered by age group [1]. Looking at crude mortality rates by age group and province, different patterns are observed. Figure 2 displays the

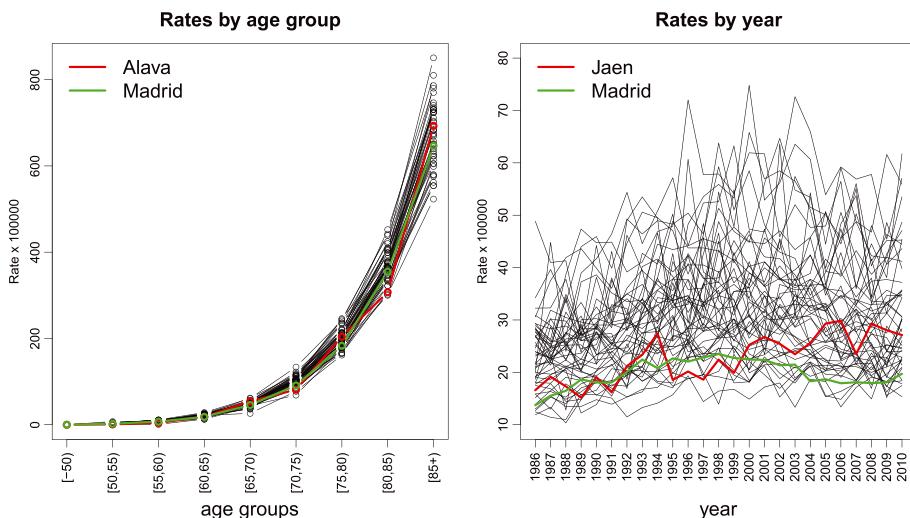


Figure 1. Evolution of prostate cancer crude mortality rates according to age (left) and according to time (right) in the 50 Spanish provinces. A total of 127,910 deaths were observed in the whole period with an overall rate of 25.5 per 100,000 male inhabitants.

Table I. Descriptive statistics by age group: total number of deaths and crude rates (expressed by 100,000 male inhabitants) for the nine age groups.

Age group	Total deaths	Crude rates
[−50)	364	0.10
[50, 55)	822	2.82
[55, 60)	2156	7.93
[60, 65)	5268	21.01
[65, 70)	10,815	49.16
[70, 75)	18,466	101.78
[75, 80)	26,446	198.46
[80, 85)	29,739	369.76
[85+)	33,834	666.87

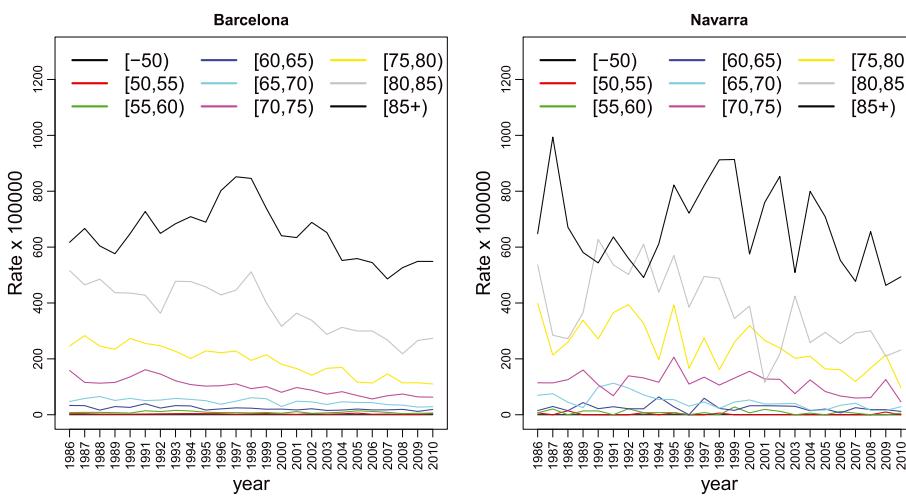


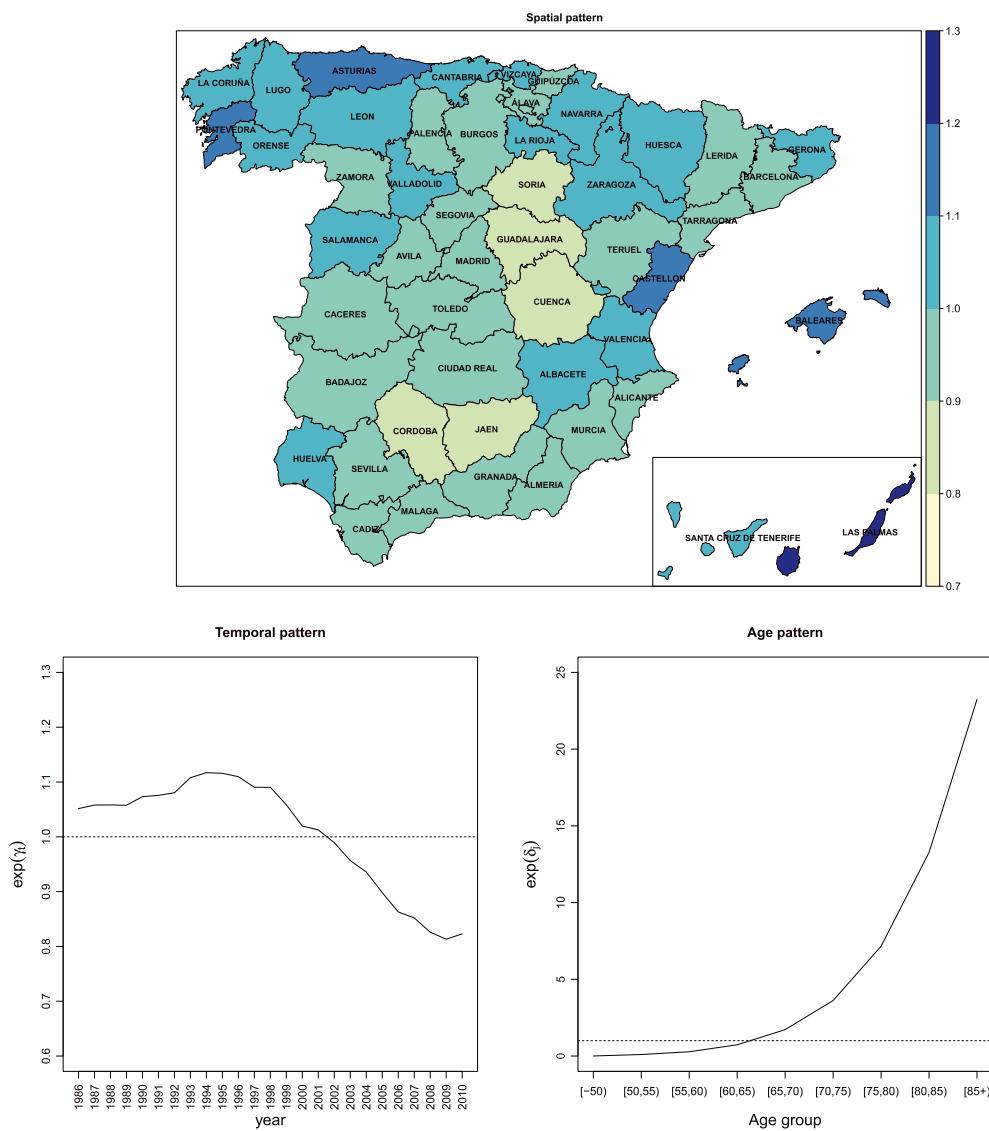
Figure 2. Temporal evolution of crude prostate cancer mortality rates (by 100,000 male inhabitants) in two Spanish provinces: Barcelona (a high populated area with a male population of 2,619,016 in 2010) and Navarre (a low populated area with a male population of 308,877 in 2010).

temporal evolution of crude mortality rates by age group in two selected provinces. One populated area, Barcelona, is displayed on the left. In this case, crude rates are rather stable, but they are different among the age groups. In low populated areas, as Navarra on the right, highly variable rates are obtained. From Figures 1 and 2 and Table I, it is clear that models to smooth rates are necessary and that the temporal evolution of rates differs among the age groups and provinces. Therefore, age–space–time models should be considered.

The first eight models described in Section 2 have been fitted to the data using a simplified Laplace strategy. Model 9 was fitted using a Gaussian approximation to reduce computational cost. Table II displays the posterior deviance (\bar{D}), the effective number of parameters p_D , the DIC , the modified DIC (DIC_C), the modified effective number of parameters p_{D_c} [31], and the logarithmic score (LS). Model 1 exhibits the highest DIC , DIC_C , and LS indicating the necessity of including interactions. The most complex models, that is Model 8 and Model 9, display the smallest quantities of posterior deviance \bar{D} , DIC , DIC_C , and LS, and consequently, they are the best candidates. Looking at the DIC_c criterion, the difference between both models is not large, and the LS is also similar for both models. It should be observed that Model 9 has been fitted using the simplest Gaussian approximation, and the computational cost is too large (nearly 6 days), so a more precise simplified Laplace strategy looks unpractical. Hence, to make a fair comparison, Model 8 was also fitted using a Gaussian strategy (lasting about 5 min), and results were very similar. Then, Model 8 was finally retained as it is simpler, but we kept the results obtained with the simplified Laplace fit (lasting about 21 h). All models were fitted in a twin superserver with four processors Intel Xeon 6C and 96GB RAM using R (version 3.2.2) and the R package INLA (INLA version

Table II. DIC , modified DIC_c , and logarithmic score values for the 9 age-space-time models

Model	\bar{D}	p_D	p_{Dc}	DIC	DIC_c	LS
Model1	47,012.92	72.47	73.36	47,085.39	47,086.28	2.094
Model2	46,880.25	121.63	123.98	47,001.88	47,004.22	2.090
Model3	46,550.16	211.28	220.23	46,761.45	46,770.40	2.080
Model4	46,360.59	109.39	111.52	46,469.98	46,472.12	2.066
Model5	46,222.43	159.73	164.04	46,382.16	46,386.47	2.062
Model6	46,418.96	259.21	272.03	46,678.17	46,690.99	2.077
Model7	45,913.44	245.20	257.34	46,158.65	46,170.79	2.052
Model8	45,774.09	294.61	311.44	46,068.70	46,085.53	2.047
Model9	45,714.65	341.66	365.55	46,056.31	46,080.20	2.047

**Figure 3.** Prostate cancer spatial effects $\exp(\phi_i)$ (top), temporal effects $\exp(\gamma_t)$ (left bottom), and age effects $\exp(\delta_j)$ (right bottom) common to all provinces. Values greater than 1 indicate that the spatial, temporal, or the age effect increase the global rate.

0.0-1440400394; INLA date Mon 24 Aug 09:13:14 CEST 2015). Once the best candidate has been chosen, it is interesting to look into the different components of the estimated mortality rates.

Figure 3 displays the global spatial pattern $\exp(\phi_i)$ in the top row, the global temporal pattern $\exp(\gamma_t)$ (left bottom row), and the global age effect $\exp(\delta_j)$ (right bottom row). The spatial pattern, where the names of the provinces have been included to make reading and interpretation easier, reveals a marked effect of Pontevedra, Asturias, Castellón, Balearic Islands, and Las Palmas (Canary Islands) on mortality rates. Note that the location of the Canary Islands is shown in an inset at the bottom right of the Spanish map. The global temporal trend reveals a global increase in rates until middle 1990s, and from then on rates go into a decline. From 1986 to 2002, the common temporal trend is greater than 1 contributing to increase the rates. On the other hand, from 2002 onwards, the temporal trend contributes to reduce the rates. The increase may be due to improvements in diagnosis and certification of deaths, whereas the decrease may be attributed to improvements in survival (see [35] for an analysis of several cancer sites in Spain by municipalities). Finally, rates increase with age, and this upward trend is steeper in the oldest age groups, something previously observed in the preliminary exploratory analysis. More precisely, the four youngest age groups have a decreasing effect on rates, whereas an increasing effect is observed for the other groups.

Model 8 also includes additional interaction terms. Figure 4 displays the area–age interaction in two alternative ways, being clear that the age groups are not equally affected by the area. The picture on the left displays the evolution of the age effect in each province. In color, four different provinces have been selected for their different behavior. On one hand, the effects of Zaragoza and Huesca, in the north of the country, increase with age. On the other hand, the effects of Badajoz and Cádiz, in the south of Spain, decrease with age. More precisely, Zaragoza and Huesca contribute to reduce the rates for the youngest five age groups and to increase the rates for the four oldest age groups. Conversely, Badajoz and Cádiz contribute to increase the rates until the age group [70, 75] and to decrease the rates for the three oldest age groups. On the right of Figure 4, area–age interactions are displayed as maps for each age group. If we focus on the same provinces, we observe that Zaragoza and Huesca (north-central part of the map) are light colored (values smaller than 1) for the five youngest age groups, and they get darker (values greater than 1) in the maps corresponding to the last four age groups. If we move to the bottom left corner of Spain, we can observe a group of provinces (with Badajoz and Cádiz among them) in dark for the first six age groups, and they get values smaller than 1 from the age group [75, 80) onward. In general, regional effects are rather similar among the three youngest age groups ($[-50], [50, 55]$, and $[55, 60]$). They are also similar among the intermediate age groups ($[60, 65], [65, 70]$, and $[70, 75]$), whereas differences are observed among the oldest age groups ($[75, 80], [80, 85]$, and $[+85]$).

Figure 5 displays two pictures of the space–time interaction term. The first picture (top) displays the temporal evolution of the regional effect. Four selected provinces are colored as they present very different behavior. The effects of Navarra (north-central Spain) and Barcelona (the north-Mediterranean area) decrease with time, whereas the effects of Jaén and Córdoba (South Spain) increase with time. The second picture (bottom) shows maps of the spatio-temporal interaction. No clear spatial pattern is observed at

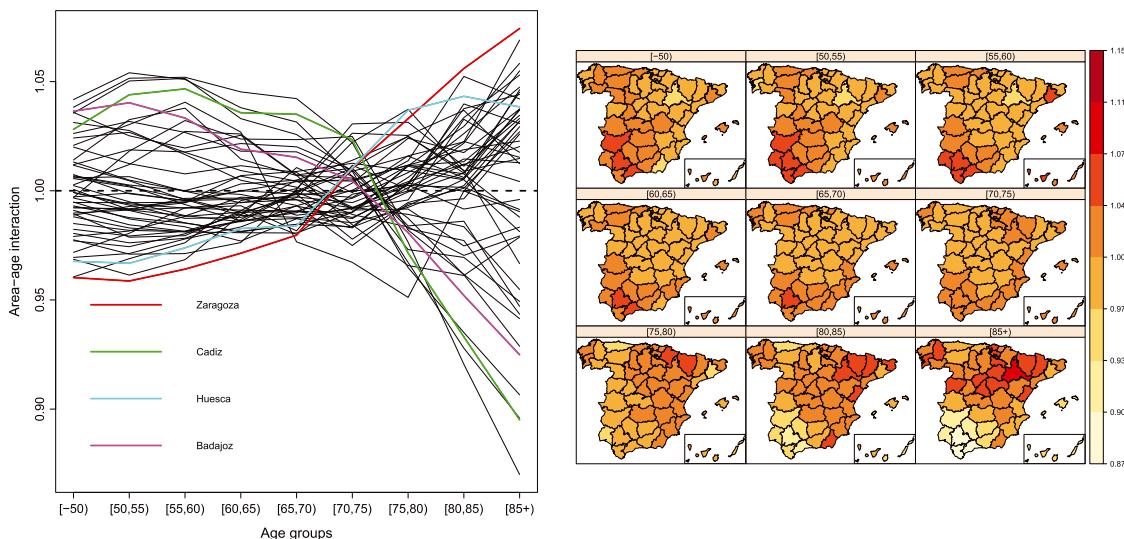


Figure 4. Area–age interaction in prostate cancer mortality. The figure on the left displays the evolution of the regional effect with age. The figure on the right displays maps of the regional effect for the different age groups.

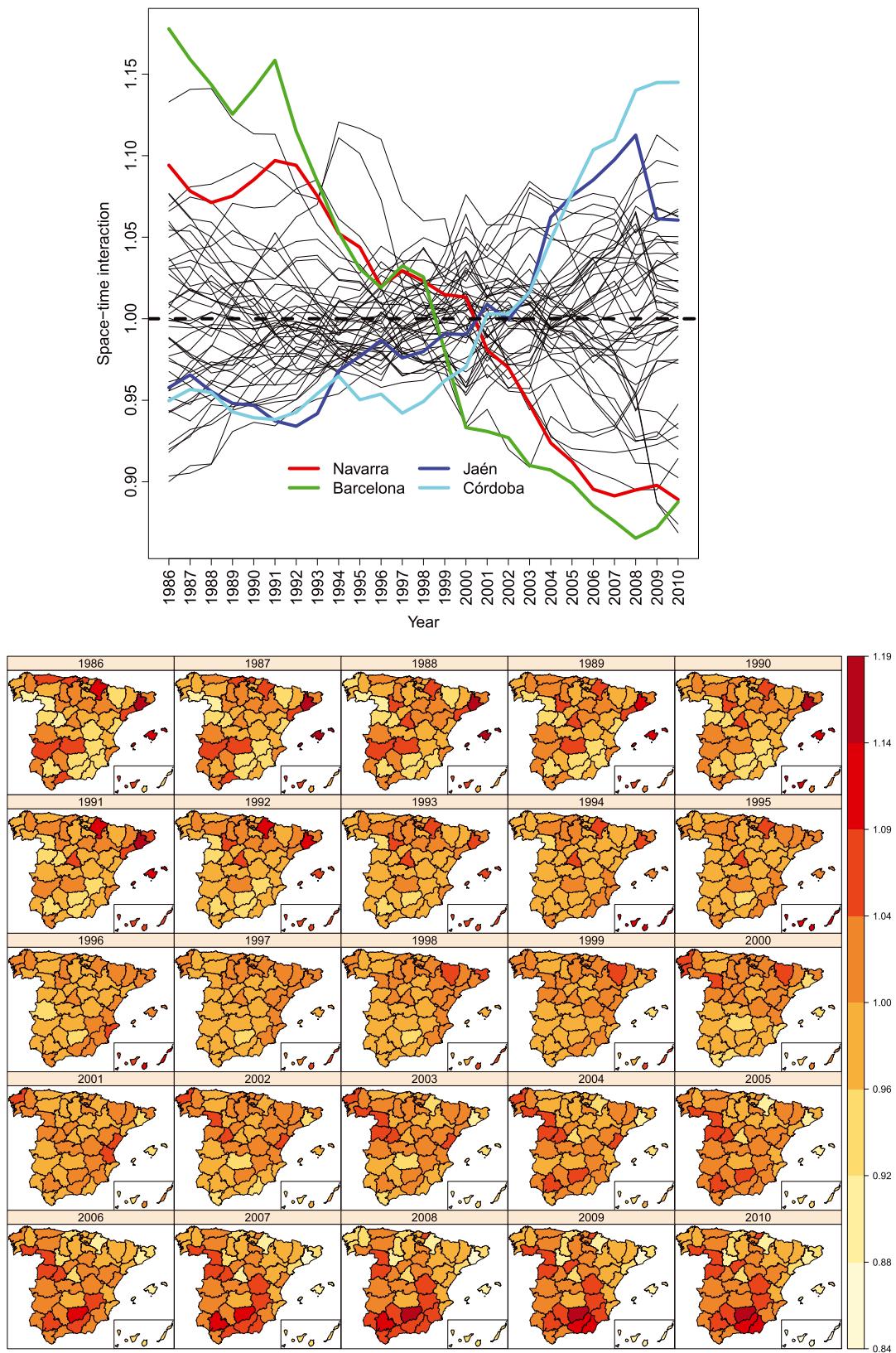


Figure 5. Space–time interaction in prostate cancer mortality. Top figure displays the temporal evolution of the regional effect. Bottom figure displays maps of the regional effects for every year of the period.

the beginning of the period, but the geographical pattern becomes homogeneous in the 1990s with the Andalusian provinces in the south having a low effect on rates. This result agrees with [35], which explain that this can be due to the protective factors of antidiabetics, since in this period, Andalusia had a higher diabetes mortality rate than the rest of Spain. However, from 2004 onwards, the effect of these regions on the rates is higher (see Jaén and Córdoba in the top Figure 5).

Figure 6 displays the temporal evolution of mortality rates (per 100,000 male inhabitants), together with credibility bands, for the different age groups in Barcelona that is a high populated area, and Navarra that is less populated. As expected, credibility intervals are narrower for high populated areas. The oldest age groups (top row) present the highest mortality rates with an increase in rates from the beginning of the period until the late 1990s. This can be due to improvements in diagnosis and certification of cause of death. Then, a decrease is observed until the end of the period that could be explained by improvements in survival. Regarding age groups under 70 years in the bottom row, a decrease is observed from the beginning to the end of the period for the age groups [60, 65) and [65, 70), and a rather flat evolution is displayed for the two youngest age groups. This is expected as the rates for these age groups are very low.

Figure 7 displays the temporal evolution of the smooth spatial patterns of mortality rates for the age group [+85], the one with the highest mortality rates. From 1986 until 1992, there is a rather homogeneous spatial distribution of mortality rates in this group. From 1993 until 1998, differences appear in the north of the country, in some Mediterranean provinces (Girona, Castellón, Valencia, and Balearic Islands) and Las Palmas (Canary Islands), where a steady increase in rates is observed. Rates start to decrease in 1999, and the mortality patterns become again homogeneous in the last 5 years.

Finally, Figure 8 displays a ranking with the five provinces with the highest mortality rates (top of the figure) and the five provinces with the smallest mortality rates (bottom of the figure) for 2010, the last year of the period, in all age groups. Each province is specifically colored, so that it is easy to see its ranking within the different age groups. The numbers in the bubbles represent the smooth rates for each province and age group. Bubbles have not been drawn with size exactly proportional to the estimated rates, but in general, a big bubble indicates a high estimated rate (the reader can appreciate that the size of the bubbles increases with the age group). It can be observed that some provinces, such as Asturias, Pontevedra, and Albacete, are ranked among the five provinces with the highest mortality rates in nearly

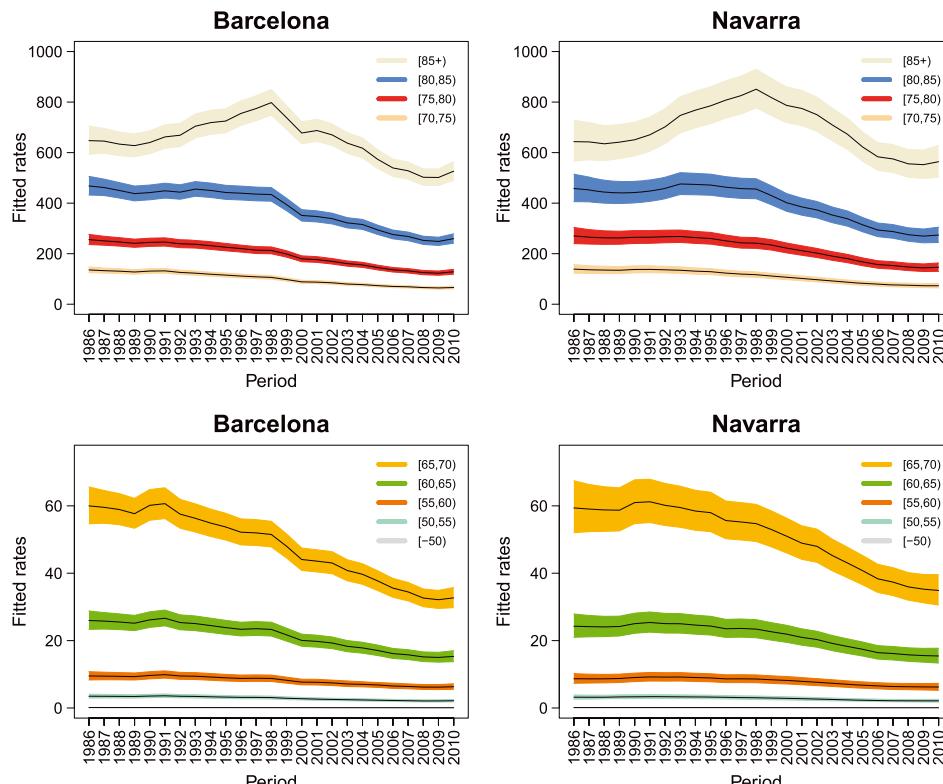


Figure 6. Temporal evolution of smooth prostate cancer mortality rates (per 100,000 male inhabitants) for the different age group in two provinces: Barcelona and Navarra.

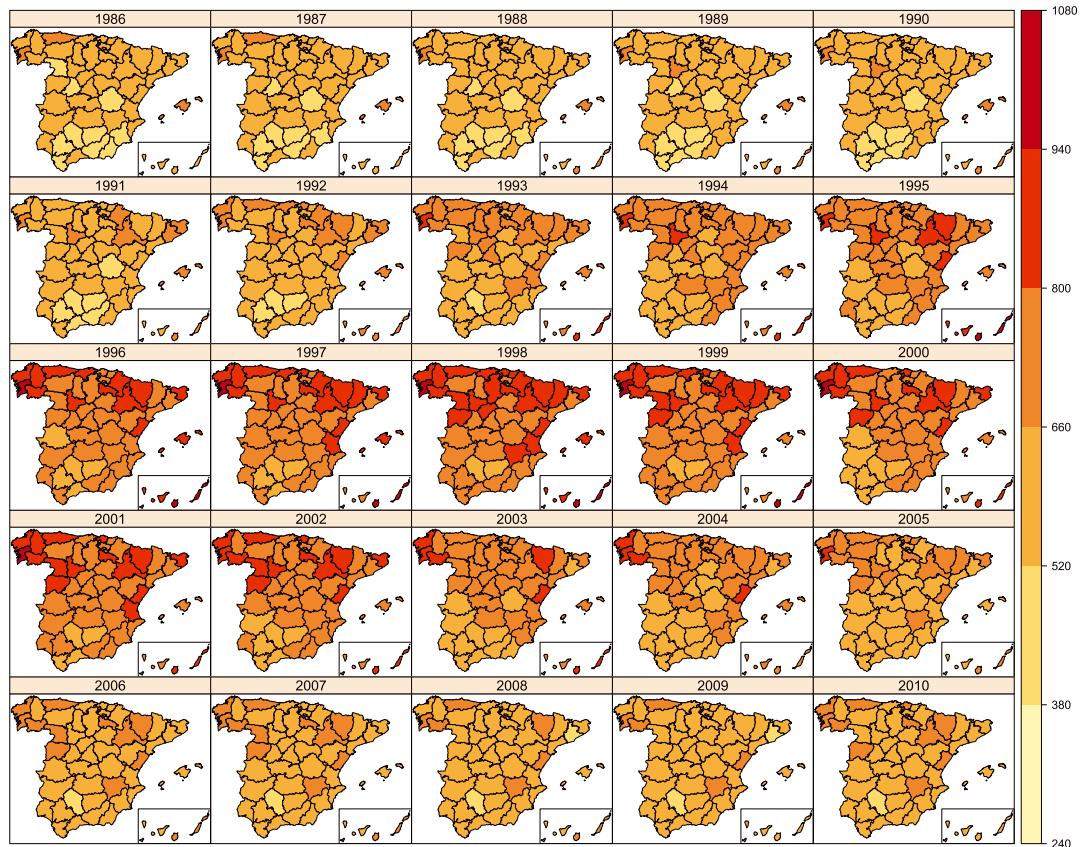


Figure 7. Temporal evolution of smooth spatial patterns of mortality rates (per 100,000 male inhabitants) for age group [+85].

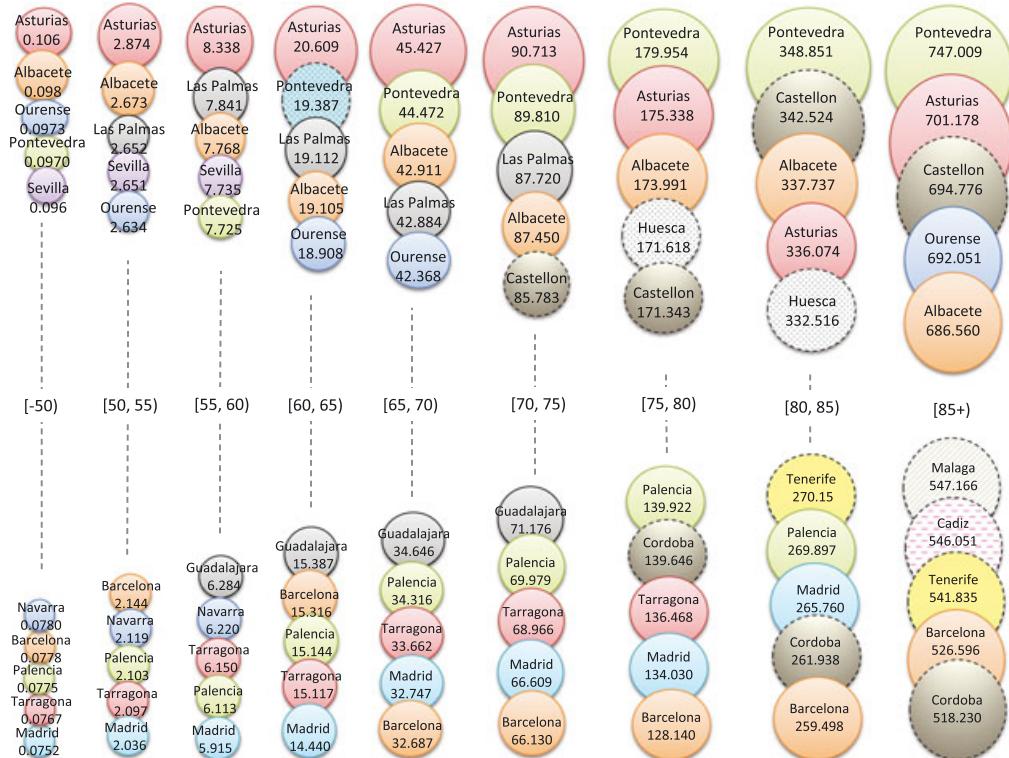


Figure 8. Ranking of regions with the highest and lowest prostate cancer mortality rates by age group for 2010, the last year of the period.

all the age groups. Other provinces such as Sevilla and Castellón are ranked among the provinces with the highest mortality rates in the youngest and oldest age groups, respectively. On the other hand, Barcelona, Madrid, Palencia, or Tarragona are ranked among the five provinces with the smallest mortality rates for almost all age groups.

5. Discussion

Aggregating counts over all age groups inside a region in a specific period of time is a common practice for studying and modeling age-standardized mortality rates. However, if a disease does not equally affect all the age groups, the age should be included in the model. Otherwise, the age effects can be masked and spatial or spatio-temporal patterns may be misconceived. Traditionally, model fitting and inference in spatio-temporal CAR models has been carried out using McMC methods, which in these models may be very time consuming as samples can be highly correlated. To overcome this difficulty, Rue et al. [15] proposed a new estimation method based on INLA. This new method is not based on simulations, and consequently, it is intended to speed up computations. The technique is particularly appealing for latent Gaussian Markov random fields whose precision matrices are sparse and depend on some hyperparameters. The models considered in this paper fit into this broad class of models. They include the Leroux et al. [16] prior for the spatial random effects, and random walks of first order for the age and time effects. One of the advantages of the INLA approach is that there is a package called R-INLA that can be used in the free software R, and therefore, practitioners have the methodology at their disposal. The spatial prior considered in this paper is not directly available in INLA yet, but it can be easily implemented as it is explained by Ugarte et al. [20]. Specific functions for random walks are available in the package.

In our real data example, a model with a triple age–space–time interaction has been discarded because of its similarity with the model including all pairwise interactions and because of its high computing time with the simplest (and less reliable) Gaussian fitting strategy. The selected model reveals interesting findings agreeing with previous results about prostate cancer. The age–time interaction shows how the region of Andalusia has a lower effect on rates during the 1990s, something that may be explained by the protective effects of antidiabetics. The great differences in rates among the age groups, particularly between the youngest and the oldest, indicate the necessity of embedding age affects. A look into the temporal trends for the age groups in different provinces shows an increase of rates in the oldest age groups until late 1990s, and from then on, rates start to decrease at different speed depending on the province. Age–space interactions also exhibit different effects of the same province on the age groups. A ranking of the provinces according to mortality rates in the last year of the period reveals that a group of provinces are always ranked among the five provinces with the highest mortality rates. All these findings may contribute to have a more complete picture of the distribution of mortality rates and to identify high rate provinces where intervention or prevention programs may be advisable.

Appendix A

In this section, the code to fit Model 8 with INLA is provided. The first step is to load to the required R packages. Next, the data file containing the variables age groups, province, year, deaths, and male population is loaded.

```
library(INLA)
library(MASS)
library(maptools)

dat<- read.table("prostate.txt", header=T)
```

The observations are setup by year, province, and age group. The order of the Kronecker products in the covariance matrices depends on how the data are arranged.

```
###Data file
#   age  prov year deaths   pop
#    1     1 1986      0 102868
#    2     1 1986      0   8296
#    3     1 1986      0   6964
#    4     1 1986      1   5504
#    5     1 1986      2   3963
```

```

#      6   1 1986     2   2996
#      7   1 1986     3   2113
#      8   1 1986     2   1199
#      9   1 1986     6    609
#
#      .   .   .     .   .
#      .   .   .     .   .
#      .   .   .     .   .
#      .   .   .     .   .
#      .   .   .     .   .
#      .   .   .     .   .
#      .   .   .     .   .
#      7   2 1986     9   3696
#      8   2 1986     7   2460
#      9   2 1986     6   1270

```

Previous to the specific INLA code, some elements are needed. Namely, the number of provinces, years, and age groups and a data frame with some indicator variables to be used in the INLA formula

```

n <- length(unique(dat$prov))           #number of provinces
t <- length(unique(dat$agno))           #number of years
e <- length(unique(dat$edad))           #number of age groups
## data.frame for the INLA formula##
DatosINLA <- data.frame(O=dat$deaths, N=dat$pop, ID.prov=dat$prov, ID.year=dat$year,
ID.age=dat$age, ID.prov.age=rep(seq(1,n*e),t), ID.prov.year=rep(seq(1,n*t),each=e),
ID.age.year=as.vector(apply(matrix(seq(1,e*t),e,t), 2, function(x) rep(x,n))))

```

where ID.prov, ID.year, and ID.age are labels for province, year, and age groups, respectively. Finally, ID.prov.age, ID.prov.year, and ID.age.year are labels for the interactions terms space-age, space-time, and age-time, respectively.

The covariance for the CAR prior for the random effect given by Leroux et al. [16] is not directly available in INLA, but it can be implemented defining the matrix in Equation (3).

```

## Spatial precision matrix (Leroux CAR) ##
g <- inla.read.graph("esp_prov_nb.inla") ## inla object with the spatial
                                         ## neighborhood structure
Qs = matrix(0, g$n, g$n)                 ## neighborhood matrix
for (i in 1:g$n){
  Qs[i,i]=g$nnbs[[i]]
  Qs[i,g$nnbs[[i]]]=-1
}
R.Leroux <- diag(dim(Qs)[1])-Qs          ## Matrix C in Equation (3)

```

The next step is to define the structure matrices for time and age and the vectors expanding the null space of the interaction.

```

Dm<-diff(diag(e),differences=1)          ## First order difference matrix for time
Qe<-t(Dm) %*% Dm                         ## Structure matrix for the temporal
                                             ## random effect
Dm<-diff(diag(t),differences=1)          ## First order difference matrix for age
Qt<-t(Dm) %*% Dm                         ## Structure matrix for the age
                                             ## random effect
null.space <- kronecker.null.space(Qt,Qs)
R.st <- null.space[[1]]                    ## Structure matrix of the
                                         ## space-time interaction
                                         ## Null space of the space-time interaction
A_delta.st <- as.matrix(null.space[[2]])    ## Structure matrix of the
null.space <- kronecker.null.space(Qs,Qe)  ## Null space of the space-time interaction
R.se <- null.space[[1]]                    ## Structure matrix of the
                                         ## space-age interaction
                                         ## Null space of the space-age interaction
A_delta.se <- as.matrix(null.space[[2]])    ## Structure matrix of the
null.space <- kronecker.null.space(Qt,Qe)  ## Null space of the space-age interaction
R.et <- null.space[[1]]                    ## Structure matrix of the
                                         ## time-age interaction
                                         ## Null space of the time-age interaction
A_delta.et <- as.matrix(null.space[[2]])    ## Null space of the time-age interaction

```

where "kronecker.null.space" is a function to compute the structure matrix of the interaction and the eigenvectors expanding the null space.

Next, we indicate how to fit Model 8 using the R-INLA package. This model is specifically described because it is the one finally chosen in the case study. INLA uses a generic function `f()` to model random effects. Here we use "`rw1`" for random walks of first order, "`generic1`" and "`generic0`" for the spatial effects and the interaction terms, respectively. Hyperpriors also specified within the `f()` function with the "`hyper`" option. The constraints are given using "`constr`" and "`extraconstr`". We first have to specify the model formula

```
formula.M8 <- O ~ f(ID.area, model="generic1", Cmatrix = R.Leroux, constr=TRUE,
                      hyper=list(prec=list(prior="loggamma", param=c(1,0.01)),
                                  beta=list(prior="logitbeta", param=c(1,1)))) +
  f(ID.year, model="rw1") +
  f(ID.edad, model="rw1") +
  f(ID.area.edad, model="generic0", Cmatrix=R.se, constr=TRUE,
  extraconstr=list(A=A_delta.se, e=rep(1e-5,dim(A_delta.se)[1]))) +
  f(ID.area.year, model="generic0", Cmatrix=R.st, constr=TRUE,
  extraconstr=list(A=A_delta.st, e=rep(1e-5,dim(A_delta.st)[1]))) +
  f(ID.edad.year, model="generic0", Cmatrix=R.et, constr=TRUE,
  extraconstr=list(A=A_delta.et, e=rep(1e-5,dim(A_delta.et)[1])))
```

where `R.Leroux` is the C matrix defined in Eq. 3, `generic0` is the INLA model for the interactions, and `A_delta.se`, `A_delta.st`, and `A_delta.et` are the A_{ζ^k} , $k = 1, 2, 3$ matrices of eigenvectors expanding the null space of the precision matrices \mathbf{Q}_{ζ^k} , $k = 1, 2, 3$. Note that "`e`" is a vector of zeros to include the constraints. Once all components of the formula are defined, the model is fitted using the function `inla()`

```
Model8 <- inla(formula.M8, family="poisson", data=DatosINLA, E=N,
                 control.predictor=list(compute=TRUE, cdf=c(log(1))),
                 control.compute=list(dic=TRUE, cpo=TRUE),
                 control.inla=list(strategy="simplified.laplace"))
```

where `family="poisson"` indicates the distribution of the response variable; `E=N` denotes population at risk for each province, age group, and year; `control.compute=list(dic=TRUE, cpo=TRUE)` is an option to compute DIC; and `control.inla=list(strategy="simplified.laplace")` specifies the approximation technique (simplified Laplace).

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Innovation (Project MTM 2011-22664 co-funded with FEDER grants), by the Spanish Ministry of Economy and Competitiveness (Project MTM 2014-51992-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. Silva GL, Dean CB. Modelling and Analysis of Disease Incidence Rates by Age Groups Over Regions. CEAUL, 2012. <http://www.ceaul.fc.ul.pt/notas.html?ano=2012> [Accessed on 15 January 2016].
2. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; **43**(1):1–20.
3. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 2000; **19**:2555–2567.
4. Ugarte MD, Goicoa T, Ibáñez B, Militino AF. Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics* 2009; **20**:647–665.
5. Ugarte MD, Goicoa T, Militino AF. Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics* 2010; **21**(3-4):270–289.
6. MacNab YC, Gustafson P. Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. *Statistics in Medicine* 2007; **26**:4455–4474.
7. Martínez-Beneito MA, López-Quílez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* 2008; **27**:2874–2889.
8. Lagazio C, Biggeri A, Dreassi E. Age-period-cohort models and disease mapping. *Environmetrics* 2003; **14**:475–490.
9. Dean CB, Ugarte MD, Militino AF. Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics* 2001; **57**(1):197–202.
10. Hoem J. Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review. *International Statistical Review* 1987; **55**:119–152.
11. Nandram B, Sedransk J, Pickle LW. Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association* 2000; **95**:1110–1118.

12. MacNab YC, Dean CB. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics* 2001; **57**:949–956.
13. Sun D, Tsutakawa RK, Kim H, He Z. Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 2000; **19**:2015–2035.
14. Zhang S, Sun D, He CZ, Schootman M. A Bayesian semi-parametric model for colorectal cancer incidences. *Statistics in Medicine* 2006; **25**:285–309.
15. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009; **71**(2):319–392.
16. Leroux BG, Lei X, 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*, Halloran M, Berry D (eds). Springer-Verlag: New York, 1999; 179–191.
17. MacNab YC. On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research* 2011; **20**(1):49–68.
18. Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC: Boca Raton, 2005.
19. Clayton DG. Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London, 1996; 275–301.
20. Ugarte MD, Adin A, Goicoa T, Militino AF. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research* 2014; **23**(6):507–530.
21. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9–25.
22. Knorr-Held L, Rue H. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics* 2002; **29**:597–614.
23. Schmid V, Held L. Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* 2004; **60**:1034–1042.
24. Schrödle B, Held L, Riebler A, Danuser J. 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)* 2011; **60**(2):261–279.
25. Rue H, Martino S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference* 2007; **137**(10):3177–3192.
26. Martino S, Rue H. Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the `inla` program. Department of Mathematical Sciences, NTNU, Norway, 2009. (Available from <http://www.math.ntnu.no/~hrue/GMRFsim/manual.pdf>) [Accessed on 15 January 2016].
27. R Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2015. <http://www.R-project.org/>.
28. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology* 2013; **4**:33–49.
29. Schrödle B, Held L. Spatio-temporal disease mapping using INLA. *Environmetrics* 2011; **22**(6):725–734.
30. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.
31. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics* 2008; **9**:523–539.
32. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 2007; **102**(477):359–378.
33. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. *GLOBOCAN (2010) Cancer Incidence and Mortality Worldwide*, IARC CancerBase N 10. Version 1.2 Lyon. IARC: France, 2008.
34. Larrañaga N, Galceran J, Ardanaz E, Franch P, Navarro C, Sánchez MJ, Pastor-Barriosuso R. Prostate cancer incidence trends in Spain before and during the prostate-specific antigen era: impact on mortality. *Annals of Oncology* 2010; **21**: 83–89.
35. López-Abente G, Aragón N, Pérez-Gómez B, Pollán M, García-Pérez J, Ramis R, Fernández-Navarro P. Time trends in municipal distribution patterns of cancer mortality in Spain. *BMC Cancer* 2014; **14**:535.