ON PREDICTING CANCER MORTALITY USING ANOVA-TYPE P-SPLINE MODELS

Authors: JAIONE ETXEBERRIA
– Department of Statistics and O. R., Public University of Navarre,
Consortium for Biomedical Research in Epidemiology and Public Health
(CIBERESP), Spain
jaione.etxeberria@unavarra.es

MARÍA DOLORES UGARTE
– Department of Statistics and O. R., Public University of Navarre,
Inst. for Advanced Materials (InaMAT), Public Univ. of Navarre, Spain
lola@unavarra.es

TOMÁS GOICOA
– Department of Statistics and O. R., Public University of Navarre,
Research Network on Health Services in Chronic Diseases (REDISSEC),
Inst. for Advanced Materials (InaMAT), Public Univ. of Navarre, Spain
tomas.goicoa@unavarra.es

ANA F. MILITINO
– Department of Statistics and O. R., Public University of Navarre,
Inst. for Advanced Materials (InaMAT), Public Univ. of Navarre, Spain
militino@unavarra.es

Abstract:

• Extrapolating cancer mortality trends can be very valuable as a tool to predict cancer
burden. National Health Agencies use different models to figure out future evolution
of cancer, but they mainly work at national level. However, developed countries are
divided into different regions with their own governments and health care systems,
and this should be taken into account. In this paper, an ANOVA-type P-spline model
is considered to predict the number of mortality cases in forthcoming years in regions
within a country. The model is very interesting as it allows to split the predictions into
components representing region-specific features and characteristics common to the
whole country. Prediction variability is also calculated to provide prediction intervals.
Real data on cancer mortality are used for illustration.

1. INTRODUCTION

Prediction of future events has always been a challenge in modern societies, and statistical methods are valuable tools to forecast outcomes in many fields of daily life. For example, in economy, we are constantly receiving predictions about employment, rate of growth, income, expenses and many other quantities. In medicine, it is common to make predictions about the evolution of a disease, the spread of an epidemic, the outbreak of influenza, or the number of new HIV cases. The main reason why governments, institutions or private companies demand predictions is that advance knowledge about the future allows to make plans, to think about business strategies and management, or to allocate resources efficiently.

Future information about cancer incidence or mortality is essential for Public Health Agencies since this illness brings huge expenses in developed countries involving diagnosis, treatment, research, loss of productivity because of sick leaves, or pensions due to premature deaths in a family. These figures are also important to efficiently organize cancer screening programs and to prioritize prevention activities. A cancer situation assessment requires an appraisal of the problem in terms of the number of incidence or mortality cases. This should be done based on an updated collection of cancer figures provided by population-based cancer registries or censuses. Regarding the official figures, these are available with a delay of approximately three or four years due to the complexity of updating cancer registries. Hence, Health Agencies substitute this lack of information with projections of cancer cases based on statistical models. Most of these agencies use models at country level, and hence, they are essentially temporal models. To show some examples, Lee *et al.* (2011) provide a comparison of the different methods using Canadian cancer mortality data for twelve cancer sites. The authors compare a temporal Poisson log-linear model used by the Public Health Agency of Canada; age-period-cohort models considered by the Association of the Nordic Cancer Registries; autoregressive with time trends models used by the American Cancer Society, or state-space models used by the National Cancer Institute. Joint point regression models implemented in the Jointpoint Regression Programme by the National Cancer Institute are also studied. According to these authors, no model can be used for all cancer sites, and the performance also depends on the number of observed cases. Moreover, the same models can show different behavior in different countries. For example, for testis, thyroid and ovary cancers, different performance is observed with Canadian and American data.

There has been additional academic research on predicting cancer mortality cases mainly based on time models. For example, Chen *et al.* (2012) and Zhu *et al.* (2012) evaluate different models to provide 4-year-ahead cancer counts projections in USA. Tiwari *et al.* (2004) consider state-space methods to improve

current projection methods used by the American Cancer Society. Ghosh and Tiwari (2007) proposed a local linear model for short term projections and a quadratic local model for longer prediction periods. Ghosh *et al.* (2008) develop a projection method based on state-space models combining the best features of a local quadratic model and an autoregressive model with fixed trend. Some more work on temporal models includes Dyba and Hakulinen (2000) or Malvezzi *et al.* (2012, 2013) to cite some of them. These models consider the calendar year as the relevant time axis. However, for some cancer sites, the relevant time axis is not the calendar year but the cohort of birth, and consequently age-period-cohort models could be used. Research about cancer projections using age-period-cohort models without spatial correlation can be found in Knorr-Held and Rainer (2001), Clements *et al.* (2005), Riebler and Held (2010) or Riebler *et al.* (2012). On the other hand, Schmid and Held (2004) provide stomach cancer mortality projections using age-period-cohort models including spatial correlation. Very recently Ugarte *et al.* (2012a) consider a three-dimensional P-spline model to project prostate cancer mortality counts in fifty Spanish provinces. The authors conclude that the P-spline model, that takes into account spatial dependencies, is preferable to individual P-spline temporal models fitted separately in each province. Etxeberria *et al.* (2014) compare different conditional autoregressive models (CAR), P-spline models, and a combination of both in terms of their predictive performance using cancer mortality data. Results reveal that models combining CAR random effects for space and P-splines for time perform slightly worse than models based only on P-splines or CAR models. The key point of these papers is that the authors provide a unified framework of smoothing and predicting under the mixed model theory using the mixed model representation of P-spline models. In a different context, Currie *et al.* (2004) use P-splines to smooth and forecast mortality rates for the pension industry, but they do not use the mixed model reformulation. In an economic setting, Ugarte *et al.* (2009) forecast dwelling prices in different neighbourhoods of Vitoria, a Spanish city.

The goal of this paper is to provide guidelines on how to extend an ANOVA-type P-spline model to predict cancer mortality counts. Recently, Ugarte *et al.* (2012b) used this model to smooth prostate cancer mortality risks in Spain. One interesting feature of this model is that it allows to split the relative risk into a smooth trend common to all regions, a smooth spatial surface constant along the time period, and a smooth interaction term representing the region-specific temporal evolution of the risk. Projections can be then decomposed into the same components. This is of great interest from an epidemiological point of view, since the decomposition of the predicted risks into these components allows to assess if the increase/decrease of those risks is mainly attributable to a common temporal behavior of all the regions or is due to an area-specific behavior during the oncoming years. This information could lead to a better organization of cancer prevention programs, open up new research lines to investigate the differences among the areas, or just help to speculate about new risk factors.

The ANOVA-type P-spline model can also be reformulated as a generalized linear mixed model where the strategy to avoid identifiability problems is very simple. In this paper, predictions of future mortality counts derived from this model are provided under the mixed model framework such that smoothing, predicting and assessing variability are jointly accomplished. The methodology will be illustrated using Spanish prostate cancer mortality data during the period 1975–2008. This will allow us to make comparisons with alternative models previously used in the literature.

The rest of the paper is laid out as follows. Section 2 describes the extension of the ANOVA-type P-spline model and how predictions are obtained. The technique is illustrated in Section 3. A validation study is presented in Section 4. Finally, the paper ends with a discussion.

---

2. TIME EXTENDED ANOVA-TYPE P-SPLINE MODEL

ANOVA decompositions of smooth functions have been already considered in the literature. See for example Gu (2002) and Belitz and Lang (2005). Recently, Wood *et al.* (2013) propose new penalties that allow ANOVA models to be fitted using existing mixed model software. In this section, a spatio-temporal ANOVA-type P-spline model with B-spline bases is considered to estimate and predict cancer mortality figures. This model was initially used by Lee and Durbán (2011) to estimate ozone levels in Europe and by Ugarte *et al.* (2012b) to smooth risk in space-time disease mapping. Different approaches using B-splines have also been considered in the disease mapping literature (see for example MacNab and Dean, 2001; MacNab and Gustafson, 2007; Silva *et al.*, 2008). In this paper we focus on extending the ANOVA-type model to estimate and predict risks jointly using a mixed model reformulation. Suppose we have a big area (*e.g.* a country) divided into smaller regions (*e.g.* provinces), for which mortality (or incidence) counts in different time points are available. Denoting the province by the subindex $s = 1, ..., S$, the time period for observed data by $t = 1, ..., T$, and conditional on the unknown relative risk r_{st}, the number of deaths C_{st} is assumed to be Poisson distributed with mean $\mu_{st} = e_{st} r_{st}$, where e_{st} is the expected number of deaths calculated on the basis that the s-th province in time t behaves as the whole country in the studied period. Then

$$(2.1) \qquad C_{st} | r_{st} \sim Poisson(\mu_{st} = e_{st} r_{st}), \qquad \log \mu_{st} = \log e_{st} + \log r_{st}.$$

In this work, our interest lies in estimating and predicting risks and counts for each province. An extension of an ANOVA-type P-spline model will be considered. The model includes additive terms for space (longitude and latitude), time, and space-time interactions, and hence the log-risk ($\log r_{st}$) is modeled as

the sum of an intercept, a smooth term for the spatial surface, a temporal smooth trend, and a smooth term for the space-time interaction.

Let us define the extended time period encompassing observed and future values. This is denoted by $t^* = 1, ..., T, T+1, T+2, ..., T+p$, where p is the number of years to predict. Log-risks for observed and predicted values are modeled as

$$(2.2) \qquad u_{st}^* = \log r_{st}^* = \delta + f_s(x_1, x_2) + f_t(t^*) + f_{st}(x_1, x_2, t^*) = \mathbf{B}^* \boldsymbol{\theta}^* .$$

The term δ is an intercept, $f_s(x_1, x_2)$ represents the smooth spatial effect constant along the period, $f_t(t^*)$ is an extended temporal trend common to all areas, and $f_{st}(x_1, x_2, t^*)$ is the extended interaction term that can be interpreted as the specific temporal trend for each area. In these expressions x_1 and x_2 are the coordinates of the centroid of the ith small area (longitude and latitude respectively), t^* is the time (for observed and predicted values), and f_i, $i = s, t, st$ are smooth functions to be estimated using P-splines with B-spline bases. $\mathbf{B}^*$ is the extended B-spline basis and $\boldsymbol{\theta}^*$ is a vector of coefficients. The matrix $\mathbf{B}^*$ is explicitly defined as

$$(2.3) \qquad \mathbf{B}^* = [\mathbf{1}_{st}^* : \mathbf{1}_t^* \otimes \mathbf{B}_s : \mathbf{B}_t^* \otimes \mathbf{1}_s : \mathbf{B}_t^* \otimes \mathbf{B}_s] ,$$

where $\mathbf{1}_{st}^*$, $\mathbf{1}_t^*$, and $\mathbf{1}_s$ are column vectors of ones of length $S \times (T+p)$, $T+p$, and S respectively. $\mathbf{B}_s = \mathbf{B}_{s_2} \square \mathbf{B}_{s_1}$ is the spatial B-spline basis defined by the row-wise ($\square$) Kronecker product (Eilers *et al.*, 2006) of the marginal basis for longitude ($\mathbf{B}_{s_1}$) and latitude ($\mathbf{B}_{s_2}$). $\mathbf{B}_t^*$ represents the extended marginal basis for time and it is a lower block-triangular partitioned matrix given by

$$(2.4) \qquad \mathbf{B}_t^* = \begin{pmatrix} \mathbf{B}_t & \mathbf{0} \\ \mathbf{B}_{t_1} & \mathbf{B}_{t_2} \end{pmatrix} .$$

In this expression, $\mathbf{B}_t$ is the time marginal basis corresponding to the observed period ($t = 1, ..., T$), and $\mathbf{B}_{t_1}$ and $\mathbf{B}_{t_2}$ are the rows corresponding to the extended data.

To ensure that f_i, $i = s, t, st$ are smooth functions, the P-spline approach places penalties on the coefficients $\boldsymbol{\theta}^*$. The extended penalty matrix $\mathbf{P}^*$ is given by a block-diagonal matrix whose components are penalties for the two-dimensional spatial component, the one dimensional time component and the three-dimensional component (space-time interactions). More precisely, $\mathbf{P}^* = \mathrm{diag}\,(\mathbf{0}, \mathbf{P}_s, \mathbf{P}_t^*, \mathbf{P}_{st}^*)$, where

$$(2.5) \quad \begin{aligned} \mathbf{P}_s &= \lambda_{s_1} \mathbf{I}_{m_2} \otimes \mathbf{P}_{s_1} + \lambda_{s_2} \mathbf{P}_{s_2} \otimes \mathbf{I}_{m_1} , \\ \mathbf{P}_t^* &= \lambda_t \mathbf{P}_{t^*} , \\ \mathbf{P}_{st}^* &= \tau_{s_1} \mathbf{I}_{m_3}^* \otimes \mathbf{I}_{m_2} \otimes \mathbf{P}_{s_1} + \tau_{s_2} \mathbf{I}_{m_3}^* \otimes \mathbf{P}_{s_2} \otimes \mathbf{I}_{m_1} + \tau_t \mathbf{P}_{t^*} \otimes \mathbf{I}_{m_2} \otimes \mathbf{I}_{m_1} . \end{aligned}$$

In these expressions, $\mathbf{I}_{m_j}$, $j = 1, 2, 3$ are identity matrices of dimension $m_j \times m_j$, where m_j is the number of columns of $\mathbf{B}_j$, $j = s_1, s_2, t$, and $\mathbf{I}_{m_3}^* = \begin{pmatrix} \mathbf{I}_{m_3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. $\mathbf{P}_{s_1}$ and $\mathbf{P}_{s_2}$ are penalty matrices for longitude and latitude respectively defined by $\mathbf{P}_{s_j} = \mathbf{D}'_{s_j} \mathbf{D}_{s_j}$, $j = 1, 2$ where $\mathbf{D}_{s_j}$ are second order difference matrices to achieve smoothness over adjacent marginal coefficients (see Eilers and Marx, 1996). Matrix $\mathbf{P}_{t^*}$ is defined using the extended difference matrix $\mathbf{D}_t^*$ for the time component given by the next expression

$$(2.6) \qquad \mathbf{D}_t^* = \begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{E}_t & \mathbf{L}_t \end{pmatrix}, \qquad \mathbf{P}_{t^*} = \mathbf{D}_t^{*'} \mathbf{D}_t^* = \begin{pmatrix} \mathbf{P}_t + \mathbf{E}'_t \mathbf{E}_t & \mathbf{E}'_t \mathbf{L}_t \\ \mathbf{L}'_t \mathbf{E}_t & \mathbf{L}'_t \mathbf{L}_t \end{pmatrix}.$$

$\mathbf{D}_t$ and $\mathbf{P}_t$ are the difference matrix and the penalty matrix for the observed time period, $\mathbf{E}_t$ and $\mathbf{L}_t$ are the rows used to obtain the penalty for the oncoming years, and λ_{s_1}, λ_{s_2}, λ_t, τ_{s_1}, τ_{s_2} and τ_t are different smoothing parameters corresponding to space, time, and interaction components respectively. The extended B-spline basis for time in Equation (2.4) and the extended difference and penalty matrices in Equation (2.6) are equal to those obtained in a three-dimensional P-spline model by Ugarte *et al.* (2012a). However, the extended transformation matrix is different. The next step is to reformulate the P-spline model (2.2) as a generalized linear mixed model. To do this, a matrix $\mathbf{T}^*$ is used to transform $\mathbf{B}^*$ into $[\mathbf{X}^* : \mathbf{Z}^*]$ and $\boldsymbol{\theta}^*$ into $(\boldsymbol{\beta}', \boldsymbol{\alpha}^{*'})'$. In this paper we provide the definition of this transformation matrix $\mathbf{T}^*$ which is based on matrices of eigenvectors corresponding to non-zero and zero eigenvalues respectively obtained from the eigen decomposition of the matrices $\mathbf{P}_i$, $i = s_1, s_2, t$. The key point in this process is to choose an extended transformation matrix preserving the original transformation matrix $\mathbf{T}$ used to fit the data. Based on the transformation matrix $\mathbf{T}$, the following extended transformation matrix is considered

$$\mathbf{T}^* = \begin{pmatrix} 1 & & & \\ & \mathbf{T}_s & & \\ & & \mathbf{T}_t^* & \\ & & & \mathbf{T}_{st}^* \end{pmatrix},$$

where $\mathbf{T}_t^*$ and $\mathbf{T}_{st}^*$ are defined by

$$\mathbf{T}_t^* = \begin{pmatrix} \mathbf{T}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_t^{-1} \end{pmatrix}, \qquad \mathbf{T}_{st}^* = \begin{pmatrix} \mathbf{T}_{st} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_t^{-1} \otimes \mathbf{I}_{m_2} \otimes \mathbf{I}_{m_1} \end{pmatrix},$$

and

$$\mathbf{T}_s = \left[1 \otimes [\mathbf{u}_{2n} \otimes \mathbf{1}_1 : \mathbf{1}_2 \otimes \mathbf{u}_{1n} : \mathbf{u}_{2n} \otimes \mathbf{u}_{1n}] : \mathbf{R}_s\right],$$

$$\mathbf{T}_t = [\mathbf{u}_{3n} \otimes 1 : \mathbf{R}_t],$$

$$\mathbf{T}_{st} = \left[\mathbf{u}_{3n} \otimes [\mathbf{u}_{2n} \otimes \mathbf{1}_1 : \mathbf{1}_2 \otimes \mathbf{u}_{1n} : \mathbf{u}_{2n} \otimes \mathbf{u}_{1n}] : \mathbf{R}_{st}\right].$$

The matrices $\mathbf{R}_s$, $\mathbf{R}_t$, and $\mathbf{R}_{st}$ are given by

$$\mathbf{R}_s = [1 \otimes [\mathbf{U}_{2s} \otimes \mathbf{U}_{1n} : \mathbf{U}_{2n} \otimes \mathbf{U}_{1s} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1s}]],$$

$$\mathbf{R}_t = [\mathbf{U}_{3s} \otimes 1],$$

$$\mathbf{R}_{st} = [\mathbf{u}_{3n} \otimes [\mathbf{U}_{2s} \otimes \mathbf{U}_{1n} : \mathbf{U}_{2n} \otimes \mathbf{U}_{1s} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1s}] :$$
$$\mathbf{U}_{3s} \otimes [\mathbf{1}_2 \otimes \mathbf{u}_{1n} : \mathbf{u}_{2n} \otimes \mathbf{1}_1 : \mathbf{u}_{2n} \otimes \mathbf{u}_{1n} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1n} : \mathbf{U}_{2n} \otimes \mathbf{U}_{1s} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1s}]].$$

Note that $\mathbf{T}_s$, $\mathbf{T}_t$ and $\mathbf{T}_{st}$ represent the components of the original transformation matrix corresponding to the observed data. $\mathbf{U}_{in} = [\mathbf{1}_i : \mathbf{u}_{in}]$ and $\mathbf{U}_{is}$, $i = 1, 2, 3$, are matrices of eigenvectors corresponding to zero and non-zero eigenvalues obtained from the eigen-decomposition of the penalty matrix $\mathbf{P}_j$, $j = s_1, s_2, t$. Using this transformation, the generalized mixed model reformulation of the extended ANOVA-type P-spline model (2.2) can be obtained. More precisely, the fixed and random effect matrices of the extended generalized linear mixed model are given by

$$\mathbf{B}^* \mathbf{T}^* = [\mathbf{1}_{st}^* : (\mathbf{1}_t^* \otimes \mathbf{B}_s)\mathbf{T}_s : (\mathbf{B}_t^* \otimes \mathbf{1}_s)\mathbf{T}_t^* : (\mathbf{B}_t^* \otimes \mathbf{B}_s)\mathbf{T}_{st}^*],$$

and the extended model is expressed as

$$(2.7) \quad \begin{pmatrix} \mathbf{u}^o \\ \mathbf{u}^p \end{pmatrix} = \delta + \begin{bmatrix} \mathbf{X}_s & \mathbf{Z}_s \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_s \\ \boldsymbol{\alpha}_s \end{pmatrix} + \begin{bmatrix} \mathbf{X}_t^o & \mathbf{Z}_t^o & \mathbf{0} \\ \mathbf{X}_t^p & \mathbf{Z}_{t_1}^p & \mathbf{Z}_{t_2}^p \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_t^p \end{pmatrix}$$
$$+ \begin{bmatrix} \mathbf{X}_{st}^o & \mathbf{Z}_{st}^o & \mathbf{0} \\ \mathbf{X}_{st}^p & \mathbf{Z}_{st_1}^p & \mathbf{Z}_{st_2}^p \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_{st} \\ \boldsymbol{\alpha}_{st} \\ \boldsymbol{\alpha}_{st}^p \end{pmatrix},$$

where detailed expressions for each of the components are given in Appendix A. Super-indexes o and p refer to matrices for observed and predicted values respectively. Note that repeated columns have been removed to avoid identifiability problems. Here $\mathbf{u}^p$ are the log-risks to be predicted; $\boldsymbol{\beta}_s$, $\boldsymbol{\beta}_t$, $\boldsymbol{\beta}_{st}$ are the fixed effects; $\boldsymbol{\alpha}_s$, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\alpha}_{st}$ are the random effects for space, time and space-time interaction respectively, corresponding to the observed data, and $\boldsymbol{\alpha}_t^p$ and $\boldsymbol{\alpha}_{st}^p$ denote random effects corresponding to predicted values.

To predict these random effects, some results on forecasting using mixed models are required, but first the covariance matrix of the random effects corresponding to the observed and predicted random effects are needed. The covariance matrix is given by $\mathbf{C} = \text{diag}(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3)$ where

$$\mathbf{C}_1 = \text{Cov}(\boldsymbol{\alpha}_s) = \mathbf{R}_s' \mathbf{P}_s \mathbf{R}_s = \begin{pmatrix} \mathbf{F}_1^{-1} & & \\ & \mathbf{F}_2^{-1} & \\ & & \mathbf{F}_3^{-1} \end{pmatrix},$$

$$\mathbf{C}_2 = \mathrm{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_t^p) = \mathbf{R}_t'^* \mathbf{P}_t^* \mathbf{R}_t^* = \begin{pmatrix} \mathbf{F}_4^{-1} & -\mathbf{F}_4^{-1}\mathbf{R}_t'\mathbf{E}_t' \\ -\mathbf{E}_t\mathbf{R}_t\mathbf{F}_4^{-1} \ \ \mathbf{I}_r + \mathbf{I}_r\mathbf{E}_t\mathbf{R}_t\mathbf{F}_4^{-1}\mathbf{R}_t'\mathbf{E}_t'\mathbf{I}_r \end{pmatrix},$$

$$\mathbf{C}_3 = \mathrm{Cov}(\boldsymbol{\alpha}_{st}, \boldsymbol{\alpha}_{st}^p) = \mathbf{R}_{st}'^* \mathbf{P}_{st}^* \mathbf{R}_{st}^*$$

$$= \begin{pmatrix} \mathbf{F}^{-1} & -\mathbf{F}^{-1}\mathbf{R}_{st}'(\mathbf{E}_t' \otimes \mathbf{I}_s) \\ -(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st}\mathbf{F}^{-1} \ \ \mathbf{I}^* + \mathbf{I}^*(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st}\mathbf{F}^{-1}\mathbf{R}_{st}'(\mathbf{E}_t' \otimes \mathbf{I}_s)\mathbf{I}^* \end{pmatrix},$$

Here, $\mathbf{R}_t^*$ and $\mathbf{R}_{st}^*$ are part of the transformation matrix $\mathbf{T}^*$ and they are given by

$$\mathbf{R}_t^* = \begin{pmatrix} \mathbf{R}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_t^{-1} \end{pmatrix}, \qquad \mathbf{R}_{st}^* = \begin{pmatrix} \mathbf{R}_{st} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_t^{-1} \otimes \mathbf{I}_{m_2} \otimes \mathbf{I}_{m_1} \end{pmatrix}.$$

Expressions for $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ and $\mathbf{F}_4$, and $\mathbf{F}$ are left out in Appendix B. Then, using these covariance matrices and the results provided by Gilmour *et al.* (2004) about prediction in mixed models, estimators for $\boldsymbol{\alpha}_t^p$ and $\boldsymbol{\alpha}_{st}^p$ are given by

$$\begin{aligned} \widehat{\boldsymbol{\alpha}_t^p} &= -\mathbf{E}_t\mathbf{R}_t\mathbf{F}_4^{-1}\mathbf{F}_4\hat{\boldsymbol{\alpha}}_t = -\mathbf{E}_t\mathbf{R}_t\widehat{\boldsymbol{\alpha}_t}, \\ \widehat{\boldsymbol{\alpha}_{st}^p} &= -(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st}\mathbf{F}^{-1}\mathbf{F}\widehat{\boldsymbol{\alpha}}_{st} = -(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st}\widehat{\boldsymbol{\alpha}_{st}}. \end{aligned}$$

(2.8)

To estimate model parameters, penalized quasi-likelihood (Breslow and Clayton, 1993) is used. The smoothing parameters become variance components, and here, the vector of variance components is $\boldsymbol{\lambda} = (\lambda_{s_1}, \lambda_{s_2}, \lambda_t, \tau_{s_1}, \tau_{s_2}, \tau_t)'$. Finally, using Equation (2.8), the estimated (corresponding to observed values) and the predicted (corresponding to future values) log-relative risks are given by

$$\begin{pmatrix} \hat{\mathbf{u}}^o \\ \hat{\mathbf{u}}^p \end{pmatrix} = \hat{\delta} + [\, \mathbf{X}_s \ \mathbf{Z}_s \,] \begin{pmatrix} \hat{\boldsymbol{\beta}}_s \\ \hat{\boldsymbol{\alpha}}_s \end{pmatrix} + \begin{bmatrix} \mathbf{X}_t^o & \mathbf{Z}_t^o \\ \mathbf{X}_t^p & \mathbf{Z}_{t_1}^p - \mathbf{Z}_{t_2}^p\mathbf{E}_t\mathbf{R}_t \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_t \\ \hat{\boldsymbol{\alpha}}_t \end{pmatrix}$$

(2.9)

$$+ \begin{bmatrix} \mathbf{X}_{st}^o & \mathbf{Z}_{st}^o \\ \mathbf{X}_{st}^p & \mathbf{Z}_{st_1}^p - \mathbf{Z}_{st_2}^p(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{st} \\ \hat{\boldsymbol{\alpha}}_{st} \end{pmatrix}.$$

3.   ILLUSTRATION

To illustrate results, Spanish prostate cancer mortality data from 1975 to 2008 are considered. This data set has been described elsewhere (see Ugarte *et al.*, 2012a, 2012b; Etxeberria *et al.*, 2014) to study different disease mapping models in terms of smoothing and prediction. We use this data set here to make comparisons with the ANOVA-type P-spline model presented in this paper. In brief, a total of 150,616 prostate cancer deaths were registered in Spain during the study period. The number of observed cases ranges from 6 to 651 depending on the province, while the number of expected cases varies from 13.76 to 794.14.

Figure 1 shows the different components of the ANOVA-type P-spline model for some Spanish provinces. Risk projections for 2009–2011 are also provided after fitting the model for the observed data using penalized quasi-likelihood. The smooth thick solid line is used for the total risk estimates and predictions, the dashed line corresponds to the temporal trend common to all areas, and the dashed-dotted line represents the area specific temporal trend. Finally, the non smooth line corresponds to the SMR's and the thin solid horizontal line is the spatial effect constant along the period. The common temporal trend is below one, and hence it contributes to decrease the mortality risk. The specific temporal trend (dashed-dotted line) can be above or below one increasing or decreasing the risk. For example, in Lugo, it is above one producing an increase in risk, even though it starts to decrease at the end of the period. It is interesting to look at Malaga or Valladolid, where the specific trend contributes to increase the risk in future years, but this is compensated for the global trend which makes the risk decrease.
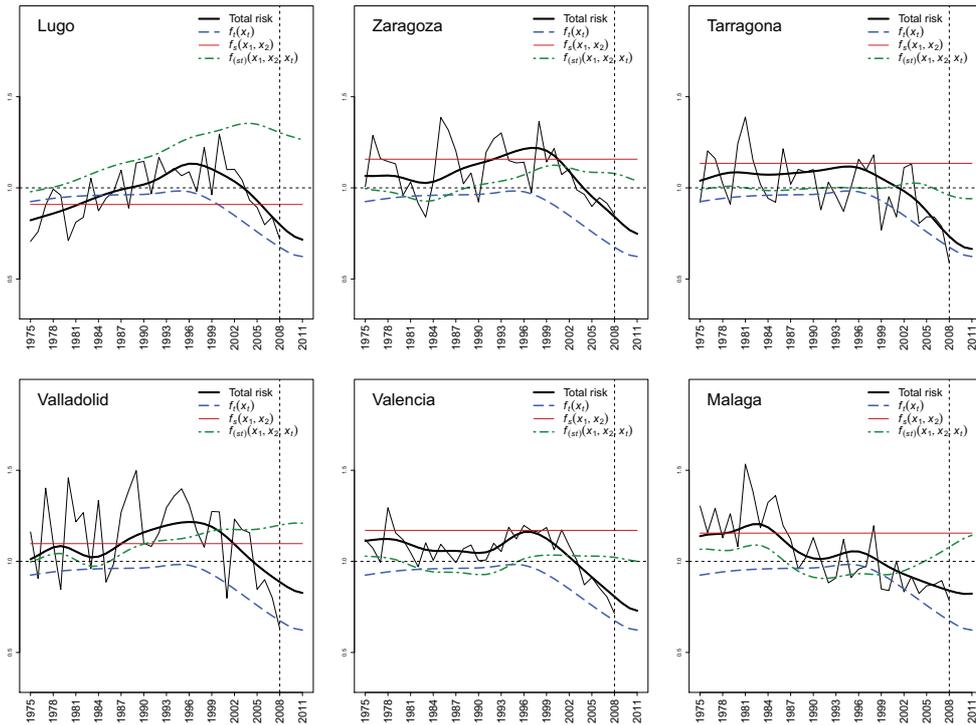
Figure 1: Risks temporal evolution of the different terms of the ANOVA-type P-spline model and predictions for the years 2009, 2010, and 2011. The smooth thick solid line corresponds to the total risk estimates and predictions, the dashed line represents the temporal trend common to all areas, and the dashed-dotted line is used for the area specific temporal trend. The non smooth line represents the SMR's, and finally the thin solid horizontal line is the spatial effect constant along the period.

Table 1:   Observed counts in 2008; risks predictions for 2011 and their corresponding 95% prediction intervals; expected counts in 2011; counts predictions for 2011 and their corresponding 95% prediction intervals.

Province	Counts 2008	Risks 2011	95% C.I. Risks 2011	Expected 2011	Counts 2011	95% C.I. Counts 2011
Coruña	161.00	0.76	[0.62, 0.93]	224.63	170.39	[123.85, 216.94]
Lleida	71.00	0.72	[0.59, 0.87]	86.24	61.73	[41.04, 82.41]
Ourense	91.00	0.76	[0.63, 0.92]	96.26	72.98	[49.72, 96.24]
Pontevedra	130.00	0.83	[0.68, 1.01]	163.42	135.11	[98.10, 172.12]
Oviedo	196.00	0.82	[0.68, 0.99]	226.20	185.59	[138.76, 232.43]
Santander	69.00	0.83	[0.69, 1.00]	107.40	89.58	[63.54, 115.62]
Lugo	53.00	0.71	[0.60, 0.86]	101.68	72.69	[49.80, 95.58]
Álava	39.00	0.73	[0.61, 0.87]	52.96	38.62	[23.97, 53.27]
Guipúzcoa	112.00	0.70	[0.58, 0.84]	125.16	87.48	[60.92, 114.05]
Vizcaya	162.00	0.78	[0.65, 0.94]	215.64	168.24	[124.91, 211.57]
Navarra	83.00	0.71	[0.60, 0.86]	112.94	80.64	[55.98, 105.29]
Huesca	54.00	0.73	[0.61, 0.88]	55.59	40.82	[25.50, 56.14]
Zaragoza	147.00	0.75	[0.62, 0.90]	183.76	137.45	[100.39, 174.51]
Teruel	31.00	0.65	[0.54, 0.78]	40.76	26.52	[14.72, 38.32]
Burgos	69.00	0.74	[0.62, 0.89]	84.56	62.84	[42.49, 83.18]
Palencia	32.00	0.79	[0.66, 0.95]	41.38	32.81	[19.68, 45.94]
León	95.00	0.76	[0.64, 0.91]	127.74	97.28	[69.36, 125.21]
Zamora	45.00	0.80	[0.66, 0.97]	60.62	48.72	[31.60, 65.84]
Valladolid	57.00	0.83	[0.69, 1.00]	98.57	81.50	[57.12, 105.88]
Soria	21.00	0.67	[0.55, 0.81]	26.60	17.72	[8.50, 26.95]
Salamanca	56.00	0.73	[0.61, 0.89]	88.01	64.57	[43.22, 85.93]
Ávila	29.00	0.70	[0.58, 0.84]	47.07	33.03	[19.59, 46.46]
Segovia	28.00	0.70	[0.59, 0.84]	38.86	27.34	[15.61, 39.06]
Logroño	62.00	0.66	[0.55, 0.79]	61.95	40.85	[25.23, 56.46]
Girona	88.00	0.72	[0.59, 0.87]	117.73	84.59	[58.33, 110.84]
Barcelona	524.00	0.64	[0.54, 0.76]	863.63	553.89	[426.37, 681.41]
Tarragona	70.00	0.66	[0.55, 0.80]	129.64	86.18	[59.70, 112.66]
Castellón	74.00	0.74	[0.62, 0.89]	94.70	69.97	[47.80, 92.14]
Valencia	253.00	0.73	[0.61, 0.87]	382.96	279.33	[213.30, 345.36]
Alicante	181.00	0.74	[0.62, 0.89]	288.08	213.82	[159.92, 267.73]
Murcia	172.00	0.73	[0.61, 0.88]	193.45	141.63	[103.02, 180.25]
Baleares	115.00	0.68	[0.55, 0.84]	145.72	99.38	[67.62 , 131.13]
Madrid	544.00	0.61	[0.52, 0.73]	871.18	533.29	[408.12, 658.47]
Cáceres	63.00	0.73	[0.60, 0.88]	85.99	62.48	[41.43, 83.53]
Badajoz	92.00	0.82	[0.68, 1.00]	119.86	98.79	[69.84, 127.74]
Guadalajara	36.00	0.57	[0.48, 0.69]	43.32	24.84	[13.41, 36.28]
Toledo	106.00	0.65	[0.55, 0.78]	121.00	78.83	[54.31, 103.34]
Cuenca	45.00	0.61	[0.51, 0.73]	54.43	33.15	[19.49, 46.81]
Ciudad Real	71.00	0.71	[0.59, 0.86]	101.42	72.14	[48.95, 95.34]
Albacete	65.00	0.77	[0.64, 0.93]	73.63	56.88	[37.86, 75.91]
Huelva	63.00	0.91	[0.74, 1.13]	71.24	65.15	[43.75, 86.54]
Sevilla	201.00	0.84	[0.69, 1.01]	240.17	201.03	[151.42, 250.63]
Cádiz	114.00	0.85	[0.70, 1.04]	149.34	126.89	[91.66, 162.12]
Córdoba	89.00	0.66	[0.55, 0.80]	130.23	86.48	[59.36, 113.60]
Málaga	147.00	0.82	[0.68, 1.00]	210.21	172.95	[128.03, 217.88]
Jaén	95.00	0.66	[0.55, 0.79]	115.29	75.98	[51.57, 100.39]
Granada	75.00	0.66	[0.54, 0.80]	138.34	91.21	[62.45, 119.97]
Almería	52.00	0.69	[0.56, 0.84]	81.31	55.83	[35.71, 75.95]
Las Palmas	113.00	0.80	[0.64, 1.01]	119.57	96.25	[65.44, 127.05]
Tenerife	110.00	0.75	[0.60, 0.94]	134.76	101.48	[68.43, 134.53]

For illustration purposes, Table 1 displays the observed counts in 2008 (the last year of the study period), risk predictions for 2011 (three year ahead predictions) together with their 95% prediction intervals; the number of expected cases for 2011 (obtained from projections of population provided by the Spanish Statistical Office), and count predictions for 2011 with their corresponding 95% prediction intervals. Confidence intervals for risks and counts are based on an appropriate estimator of the mean squared error (MSE). Traditionally, the variability associated to the estimation of the variance components have been ignored in empirical Bayes disease mapping and hence, the MSE was underestimated. This can be particularly relevant for the ANOVA-type P-spline model considered here as six smoothing parameters (variance components) are involved. The MSE for the log-risks corresponding to observed data has been derived in a spatial context by Ugarte *et al.* (2008), Escaramís *et al.* (2008) and Goicoa *et al.* (2012), and in a spatio-temporal context by Ugarte *et al.* (2010) and Ugarte *et al.* (2012b) when considering CAR, P-splines and ANOVA-type P-spline models. The MSE for predicted log-risks has also been obtained for an interaction P-spline model (Ugarte *et al.*, 2012a), and for CAR and mixtures of CAR and P-spline models (Etxeberria *et al.*, 2014). Using similar tools, the MSE estimator for projections of log-risks derived from the ANOVA-type P-spline model is computed here. The empirical coverage of confidence intervals based on this estimator reveals a good performance. To facilitate the reading of the paper technical details are given in Appendix C.

4.   VALIDATION

To assess the predicted ability of the model, a validation study is conducted. We consider the period 1995–2008 to compare the observed with the predicted counts. In brief, data from 1975–1992 are used to fit the model and to predict counts for 1995. Using data till 1993, we forecast counts for 1996 and so on. Three year ahead predictions are considered as this is normally the delay in the registers. Hence, observed counts and three-year ahead predictions from 1995 till 2008 are compared. In this validation period, predictions for 2006 and 2007 were excluded due to computational instabilities in the variance component estimates. Additionally, the models described in Etxeberria *et al.* (2014) are taken into account for comparison purposes. Namely, an additive model with a CAR structure for space and a random walk of order 2 (RW2) for time; two models with the same structure for space and time and structured and unstructured interactions (Knorr-Held, 2000); an additive model with a CAR structure for space and a P-spline for time; the same additive models with space-time interactions; a model with a common P-spline for time and specific P-splines to describe the temporal evolution of each region, and finally a pure interaction P-spline model. To make the comparison with the ANOVA-type P-spline model fair, predictions for 2006 and 2007 have been also excluded in the previous models.

Table 2 displays empirical coverage rates for prediction intervals corresponding to three, two and one year ahead predictions at nominal values 95%, and 99%. The ANOVA-type P-spline model achieves the nominal values for three year-ahead predictions, the most interesting case from a practical point of view as it is the usual delay in mortality registers. For two and one year ahead predictions, the last three models in the table (all based on P-splines) seem to attain empirical coverage rates closer to the nominal ones. The ANOVA-type P-spline model offers great flexibility and it allows to explicitly split the predictions into components representing region-specific features and characteristics common to the whole country.

Table 2:   A comparison of empirical coverage probabilities in the period 1995–2008 (excluding 2006–2007).

	Three year-ahead predictions		Two year-ahead predictions		One year-ahead predictions	
	95%	99%	95%	99%	95%	99%
Additive CAR RW2	87.67	95.00	85.00	95.00	86.50	95.50
Interaction CAR RW2 (struc.)	94.33	98.83	93.00	97.50	91.17	96.83
Interaction CAR RW2(unstruc.)	93.83	99.17	92.83	97.50	90.67	97.00
CAR(s)+Pspline(t)	85.83	94.67	87.50	94.00	88.33	95.00
CAR(s)+Pspline(t) + Int	85.50	93.67	89.33	95.50	91.33	97.50
Pspline(t) + Pspline Int	94.83	98.17	93.17	97.67	93.33	98.17
Pure interaction Pspline	93.33	99.17	93.00	98.17	91.00	97.00
ANOVA-type Pspline	95.33	99.00	92.33	97.66	91.83	97.33

5.   DISCUSSION

Statistical methods represent a valid scientific tool to make predictions about future events taking into account past information. These statistical methods gain importance in an epidemiological context since official cancer death figures are available after approximately three years from current date due to the delay in administrative procedures of data collection and registration.

Some models including CAR, P-splines and combinations of both have been studied in the literature (see for example Etxeberria *et al.*, 2014) to provide predictions of mortality or incidence counts. In this paper, an ANOVA-type P-spline model is studied to complete the P-spline alternatives within a generalized linear mixed model framework. An extended transformation matrix, including the spatial and temporal additive terms, and the spatio-temporal interaction is derived in order to express risks related to observed and future time periods in a single mixed model. The MSE of the predicted log-risks is also provided accounting

for all sources of variability, including the one coming from the estimation of the smoothing parameters, and is used in turn to calculate the count prediction error. The model has good empirical coverage rates for three year ahead predictions and in addition, it is very attractive as it explicitly considers one smooth term for space, another one for time, and a final interaction term, each one with its own smoothing parameters. This allows to split the predicted risk into a spatial component constant along the time period, a smooth temporal term common to all regions and an area specific term representing the specificity of a region. This is of practical interest as the area specific term indicates whether the region contributes to increase or decrease its own risk, and hence it helps to plan prevention or intervention measures and epidemiological policies in general.

A.   APPENDIX

To understand how the extended mixed model (2.7) is obtained, detailed expressions for the different matrices are provided in this section. Using the transformation matrix $\mathbf{T}^*$, the fixed and random effect matrices of the extended generalized linear mixed model are given by

$$\mathbf{B}^*\mathbf{T}^* = [\mathbf{1}^*_{st} : (\mathbf{1}^*_t \otimes \mathbf{B}_s)\mathbf{T}_s : (\mathbf{B}^*_t \otimes \mathbf{1}_s)\mathbf{T}^*_t : (\mathbf{B}^*_t \otimes \mathbf{B}_s)\mathbf{T}^*_{st}] \ ,$$

where

$$
\begin{aligned}
(\mathbf{1}^*_t \otimes \mathbf{B}_s)\mathbf{T}_s \quad &= \quad \mathbf{1}^*_t \otimes [\mathbf{x}_s : [\mathbf{Z}_2 \square \mathbf{X}_1 : \mathbf{X}_2 \square \mathbf{Z}_1 : \mathbf{Z}_2 \square \mathbf{Z}_1]] = \mathbf{1}^*_t \otimes [\mathbf{x}_s : \mathbf{Z}_a] \\[4pt]
&= \quad [(\mathbf{1}^*_t \otimes \mathbf{x}_s) : (\mathbf{1}^*_t \otimes \mathbf{Z}_a)] = [\mathbf{X}_s : \mathbf{Z}_s], \\[6pt]
(\mathbf{B}^*_t \otimes \mathbf{1}_s)\mathbf{T}^*_t \quad &= \quad
\begin{bmatrix} \mathbf{B}_t\mathbf{u}_{3n} & \mathbf{B}_t\mathbf{U}_{3s} & \mathbf{0} \\ \mathbf{B}_{t_1}\mathbf{u}_{3n} & \mathbf{B}_{t_1}\mathbf{U}_{3s} & \mathbf{B}_{t_2}\mathbf{L}_t^{-1} \end{bmatrix} \otimes \mathbf{1}_s =
\begin{bmatrix} \mathbf{x}^o_t \otimes \mathbf{1}_s & \mathbf{Z}^o_t & \mathbf{0} \\ \mathbf{x}^p_t \otimes \mathbf{1}_s & \mathbf{Z}^p_{t_1} & \mathbf{Z}^p_{t_2} \end{bmatrix} \\[6pt]
&= \quad
\begin{bmatrix} \mathbf{X}^o_t & \mathbf{Z}^o_t & \mathbf{0} \\ \mathbf{X}^p_t & \mathbf{Z}^{p1}_t & \mathbf{Z}^{p2}_t \end{bmatrix}, \\[6pt]
(\mathbf{B}^*_t \otimes \mathbf{B}_s)\mathbf{T}^*_{st} \quad &= \quad
\begin{bmatrix} \mathbf{x}^o_t \otimes \mathbf{x}_s : [\mathbf{x}^o_t \otimes \mathbf{Z}_a : \mathbf{B}_t\mathbf{U}_{3s} \otimes \mathbf{x}_s : \mathbf{B}_t\mathbf{U}_{3s} \otimes \mathbf{Z}_a] : & \mathbf{0} \\ \mathbf{x}^p_t \otimes \mathbf{x}_s : [\mathbf{x}^p_t \otimes \mathbf{Z}_a : \mathbf{B}_{t_1}\mathbf{U}_{3s} \otimes \mathbf{x}_s : \mathbf{B}_{t_1}\mathbf{U}_{3s} \otimes \mathbf{Z}_a] : & (\mathbf{B}_{t_2}\mathbf{L}_t^{-1}) \otimes \mathbf{B}_s \end{bmatrix} \\[6pt]
&= \quad
\begin{bmatrix} \mathbf{X}^o_{st} & \mathbf{Z}^o_{st} & \mathbf{0} \\ \mathbf{X}^p_{st} & \mathbf{Z}^p_{st_1} & \mathbf{Z}^p_{st_2} \end{bmatrix}.
\end{aligned}
$$

Here, $\quad \mathbf{Z}_s = (\mathbf{1}^*_t \otimes \mathbf{Z}_a), \quad \mathbf{Z}_a = [\mathbf{Z}_2 \square \mathbf{X}_1 : \mathbf{X}_2 \square \mathbf{Z}_1 : \mathbf{Z}_2 \square \mathbf{Z}_1], \quad \mathbf{X}_s = (\mathbf{1}^*_t \otimes \mathbf{x}_s),$ $\mathbf{x}_s = [\mathbf{1}_n \square \mathbf{x}_1 : \mathbf{x}_2 \square \mathbf{1}_n : \mathbf{x}_2 \square \mathbf{x}_1], \ \mathbf{X}_1 = [1 : \mathbf{x}_1], \ \mathbf{X}_2 = [1 : \mathbf{x}_2], \ \mathbf{Z}_1 = \mathbf{B}_{s_1}\mathbf{U}_{1s}, \ \mathbf{Z}_2 = \mathbf{B}_{s_2}\mathbf{U}_{2s}$ and $\mathbf{Z}_3 = \mathbf{B}_t\mathbf{U}_{3s}$. Finally, $\mathbf{x}_1$ and $\mathbf{x}_2$ are column vectors of longitude and latitude respectively, and $\mathbf{x}^o_t$ and $\mathbf{x}^p_t$ are column vector of time corresponding to observed and prediction period respectively. Using these results, the extended model is (2.7) is attained.

B.   APPENDIX

In this section, and to make the reading easier, expressions for matrices $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, $\mathbf{F}_4$, and $\mathbf{F} = \text{blockdiag}(\mathbf{F}_5, \mathbf{F}_6, \mathbf{F}_7, \mathbf{F}_8, \mathbf{F}_9, \mathbf{F}_{10}, \mathbf{F}_{11})$ are given. These matrices are different blocks of the covariance matrix of the random effects coming from the mixed model representation of the ANOVA-type P-spline model. Note that $\mathbf{F}_i$, $i = 1, ..., 8$ are exactly the same as those in Ugarte *et al.* (2012b). $\mathbf{F}_9$, $\mathbf{F}_{10}$, $\mathbf{F}_{11}$ are not the same because in this paper we have considered $\mathbf{B}_t^* \otimes \mathbf{B}_s$, the last term in the extended basis (2.3), instead of the other way around $\mathbf{B}_s \otimes \mathbf{B}_t^*$. This has been done because it is more natural and convenient when extending the time basis to make predictions. Expressions for these matrices are given by

$$\mathbf{F}_1 \;\; = \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_2, \quad \mathbf{F}_2 = \lambda_1 \mathbf{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1, \quad \mathbf{F}_3 = \lambda_1 \mathbf{I}_{m_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \lambda_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_{m_1-2},$$

$$\mathbf{F}_4 \;\; = \lambda_t \widetilde{\boldsymbol{\Sigma}}_3, \quad \mathbf{F}_5 = \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_2, \quad \mathbf{F}_6 = \tau_1 \mathbf{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1,$$

$$\mathbf{F}_7 \;\; = \tau_1 \mathbf{I}_{m_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \tau_2 \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_{m_1-2}, \quad \mathbf{F}_8 = \tau_t \mathbf{I}_3 \otimes \widetilde{\boldsymbol{\Sigma}}_3,$$

$$\mathbf{F}_9 \;\; = \tau_2 \mathbf{I}_{m_3-2} \otimes \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_2 + \tau_t \widetilde{\boldsymbol{\Sigma}}_3 \otimes \mathbf{I}_{m_2-2} \otimes \mathbf{I}_2,$$

$$\mathbf{F}_{10} = \tau_1 \mathbf{I}_{m_3-2} \otimes \mathbf{I}_2 \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \tau_t \widetilde{\boldsymbol{\Sigma}}_3 \otimes \mathbf{I}_2 \otimes \mathbf{I}_{m_1-2},$$

$$\mathbf{F}_{11} = \tau_1 \mathbf{I}_{m_3-2} \otimes \mathbf{I}_{m_2-2} \otimes \widetilde{\boldsymbol{\Sigma}}_1 + \tau_2 \mathbf{I}_{m_3-2} \otimes \widetilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_{m_1-2} + \tau_t \widetilde{\boldsymbol{\Sigma}}_3 \otimes \mathbf{I}_{m_2-2} \otimes \mathbf{I}_{m_1-2}.$$

where $\widetilde{\boldsymbol{\Sigma}}_i$, $i = 1, 2, 3$ are diagonal matrices of non zero eigenvalues coming from the eigen-decomposition of the marginal penalties $\mathbf{P}_{s_1}$, $\mathbf{P}_{s_2}$ and $\mathbf{P}_t$ respectively. $\mathbf{F}^{-1} = blockdiag(\mathbf{F}_5^{-1}, \mathbf{F}_6^{-1}, \mathbf{F}_7^{-1}, \mathbf{F}_8^{-1}, \mathbf{F}_9^{-1}, \mathbf{F}_{10}^{-1}, \mathbf{F}_{11}^{-1})$, $\mathbf{I}_s = \mathbf{I}_{m_2} \otimes \mathbf{I}_{m_1}$, $\mathbf{I}^* = \mathbf{I}_r \otimes \mathbf{I}_s$, and $\mathbf{I}_r$ is the identity matrix of dimension $r \times r$ where r is the number of columns of $\mathbf{L}_t$.

C.   APPENDIX

The MSE for predicted log-risk has already been proposed for a three-dimensional P-spline model (Ugarte *et al.*, 2012a). Here we reproduce the expressions and make explicit the specific formula for the $\mathbf{M}$ matrix in the ANOVA-type P-spline model. An estimator for the MSE of the predicted log-risk is given by

$$\widehat{MSE}[\hat{u}_{st}^p] = g_{1st}^*(\hat{\boldsymbol{\lambda}}) + g_{2st}^*(\hat{\boldsymbol{\lambda}}) + 2g_{3st}^*(\hat{\boldsymbol{\lambda}}) \; .$$

where

$$g_{1st}^*(\boldsymbol{\lambda}) \; = \; \mathbf{z}_{st}^p(\mathbf{C} - \mathbf{M}\mathbf{Z}^{o'}\mathbf{V}^{-1}\mathbf{Z}^o\mathbf{M}')\mathbf{z}_{st}^{p\,'} \; ,$$

$$g_{2st}^*(\boldsymbol{\lambda}) \; = \; (\mathbf{x}_{st}^p - \mathbf{z}_{st}^p\mathbf{M}\mathbf{Z}^{o'}\mathbf{V}^{-1}\mathbf{X}^o)(\mathbf{X}^{o'}\mathbf{V}^{-1}\mathbf{X}^o)^{-1}(\mathbf{x}_{st}^p - \mathbf{z}_{st}^p\mathbf{M}\mathbf{Z}^{o'}\mathbf{V}^{-1}\mathbf{X}^o)' \; ,$$

$$g_{3st}^*(\boldsymbol{\lambda}) \; = \; \mathrm{tr}[\mathbf{S}^*\mathbf{V}\mathbf{S}^{*'}\mathcal{I}^{-1}] \; .$$

Here $\mathbf{V}$ and $\mathcal{I}^{-1}$ are the covariance matrix of the working vector and the asymptotic covariance matrix of the variance components estimators arising from the PQL algorithm. Vectors $\mathbf{z}_{st}^p$ and $\mathbf{x}_{st}^p$ are the st row of the matrices $\mathbf{Z}^p = [\mathbf{Z}_s : \mathbf{Z}_{t_1}^p : \mathbf{Z}_{t_2}^p : \mathbf{Z}_{st_1}^p : \mathbf{Z}_{st_2}^p]$ and $\mathbf{X}^p = [\mathbf{X}_s : \mathbf{X}_t^p : \mathbf{X}_{st}^p]$ respectively, and finally, $\mathbf{Z}^o = [\mathbf{Z}_s : \mathbf{Z}_t : \mathbf{Z}_{st}]$ and $\mathbf{X}^o = [\mathbf{X}_s : \mathbf{X}_t^o : \mathbf{X}_{st}^o]$. An explicit expression for $\mathbf{M}$ is given by

$$\mathbf{M} = \begin{pmatrix} \mathbf{C}_1 & & \\ & \begin{pmatrix} \mathbf{F}_4^{-1} \\ -\mathbf{E}_t\mathbf{R}_t\mathbf{F}_4^{-1} \end{pmatrix} & \\ & & \begin{pmatrix} \mathbf{F}^{-1} \\ -(\mathbf{E}_t \otimes \mathbf{I}_s)\mathbf{R}_{st}\mathbf{F}^{-1} \end{pmatrix} \end{pmatrix} .$$

If λ_j denotes the jth entry of the vector of variance components $\boldsymbol{\lambda} = (\lambda_{s_1}, \lambda_{s_2}, \lambda_t, \tau_{s_1}, \tau_{s_2}, \tau_t)'$, the matrix $\mathbf{S}^*$ is given by

$$\mathbf{S}_j^* = \mathbf{z}_{st}^* \left(\frac{\partial \mathbf{M}}{\partial \lambda_j} \mathbf{Z}'\mathbf{V}^{-1} + \mathbf{M}\mathbf{Z}'\frac{\partial \mathbf{V}^{-1}}{\partial \lambda_j} \right) , \qquad j = 1, 2, 3, 4, 5, 6 \, .$$

Finally, the variance for predicted counts is calculated as

$$\mathrm{Var}[C_{st}^p] = \mathrm{E}[\mathrm{Var}[C_{st}^p|r_{st}^p]] + \mathrm{Var}[\mathrm{E}[C_{st}^p|r_{st}^p]] = e_{st}^p\mathrm{E}[r_{st}^p] + e_{st}^{p2}\mathrm{Var}[r_{st}^p] \; ,$$

where e_{st}^p are projections of the number of expected cases for future years. $\mathrm{Var}[r_{st}^p]$ is easily estimated from $\widehat{MSE}[\hat{u}_{st}^p]$ using the delta method.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Science and Innovation (project MTM 2011-22664 co-funded with FEDER grants) 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]   BELITZ, C. and LANG, S. (2005). Simultaneous selection of variables and smoothing parameters in structured additive regression models, *Computational Statistics and Data Analysis*, **53**, 61–81.

[2]   BRESLOW, N.E. and CLAYTON, D.G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.

[3]   CHEN, H.S.; PORTIER, K.; GHOSH, K.; NAISHADHAM, D.; KIM H.J.; ZHU, L.; PICKLE, L.W.; KRAPCHO, M.; SCOPPA, S.; JEMAL, A. and FEUER, E.J. (2012). Predicting US- and state-level cancer counts for the current calendar year. Part I: Evaluation of temporal projection methods for mortality, *Cancer*, **118**, 1091–1099.

[4]   CLEMENTS, M.S.; ARMSTRONG, B.K. and MOOLGAVKAR, S.H. (2005). Lung cancer rate predictions using generalized additive models, *Biostatistics*, **6**, 576–589.

[5]   CURRIE, I.D.; DURBÁN, M. and EILERS, P.H.C. (2004). Smoothing and forecasting mortality rates, *Statistical Modelling*, **4**, 279–298.

[6]   DYBA, T. and HAKULINEN, T. (2000). Comparison of different approaches to incidence prediction based on simple interpolation techniques, *Statistics in Medicine*, **19**, 1741–1752.

[7]   EILERS, P.H.C.; CURRIE, I.D. and DURBÁN, M. (2006). Fast and compact smoothing on large multidimensional grids, *Computational Statistics and Data Analysis*, **50**, 61–76.

[8]   EILERS, P.H.C. and MARX, B.D. (1996). Flexible smoothing with B-splines and penalties, *Statistical Science*, **11**, 89–121.

[9]   ESCARAMÍS, G.; CARRASCO, J.L. and ASCASO, C. (2008). Detection of significant disease risks using a spatial conditional autoregressive model, *Biometrics*, **64**, 1043–1053.

[10]   ETXEBERRIA, J.; GOICOA, T.; UGARTE, M.D. and MILITINO, A.F. (2014). Evaluating space-time models for short-term cancer mortality risk predictions in small areas, *Biometrical Journal*, **56**, 383–402.

[11] GHOSH, K. and TIWARI, R.C. (2007). Prediction of U.S. cancer mortality counts using semiparametric Bayesian techniques, *Journal of the American Statistical Association*, **102**, 7–15.

[12] GHOSH, K.; TIWARI, R.C.; FEUER, E.J.; CRONIN, K. and JEMAL, A. (2008). *Predicting US cancer mortality using state-space models.* In "Computational Methods in Biomedical Research" (R. Khattree and D.N. Naik, Eds.), Chapman and Hall/CRC, Boca Raton, 131–151.

[13] GILMOUR, A.; CULLIS, B.; WELHAM, S.; GOGEL, B. and THOMPSON, R. (2004). An efficient computing strategy for prediction in mixed linear models, *Computational Statistics and Data Analysis*, **44**, 571–586.

[14] GOICOA, T.; UGARTE, M.D.; ETXEBERRIA, J. and MILITINO, A.F. (2012). Comparing CAR and P-spline models in spatial disease mapping, *Environmental and Ecological Statistics*, **19**, 573–599.

[15] GU, C. (2002). *Smoothing Spline ANOVA Models.* Springer, New York.

[16] KNORR-HELD, L. (2000). Bayesian modelling of inseparable space-time variation in disease riskm *Statistics in Medicine*, **19**, 2555–2567.

[17] KNORR-HELD, L. and RAINER, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction, *Biostatistics*, **2**, 109–129.

[18] LEE, T.C.K.; DEAN, C.B. and SEMENCIW, R. (2011). Short-term cancer mortality projections: A comparative study of prediction methods, *Statistics in Medicine*, **30**, 3387–3402.

[19] LEE, D.J. and DURBÁN, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing, *Statistical Modelling*, **11**, 49–69.

[20] MACNAB, Y.C. and DEAN, C.B. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates, *Biometrics*, **57**, 949–956.

[21] MACNAB, Y.C. and GUSTAFSON, P. (2007). Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance, *Statistics in Medicine*, **26**, 4455–4474.

[22] MALVEZZI, M.; BERTUCCIO, P.; LEVI, F.; LA VECCHIA, C. and NEGRI, E. (2012). European cancer mortality predictions for the year 2012, *Annals of Oncology*, **23**, 1044–1052.

[23] MALVEZZI, M.; BERTUCCIO, P.; LEVI, F.; LA VECCHIA, C. and NEGRI, E. (2013). European cancer mortality predictions for the year 2013, *Annals of Oncology*, **24**, 792–800.

[24] RIEBLER, A. and HELD, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Biostatistics*, **11**, 57–69.

[25] RIEBLER, A.; HELD, L. and RUE, H. (2012). Estimation and extrapolation of time trends in registry data-Borrowing strength from related populations, *The Annals of Applied Statistics*, **6**, 304–333.

[26] SCHMID, V. and HELD, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data, *Biometrics*, **60**, 1034–1042.

[27] SILVA, G.L.; DEAN, C.B.; NIYONSENGA, T. and VANASSE, A. (2008). Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines, *Statistics in Medicine*, **27**, 2381–2401

[28]   Tiwari, R.C.; Ghosh, K.; Jemalm A.; Hachey, M.; Ward, E.; Thun, M.J. and Feuer, E.J. (2004). A new method of predicting U.S. and state-level cancer mortality counts for the current calendar year, *CA: A Cancer Journal for Clinicians*, **54**, 30–40.

[29]   Ugarte, M.D.; Goicoa, T.; Etxeberria, J. and Militino, A.F. (2012a). Projections of cancer mortality risks using spatio-temporal Pspline models, *Statistical Methods in Medical Research*, **21**, 545–560.

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

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

[32]   Ugarte, M.D.; Goicoa, T.; Militino, A.F. and Durbán, M. (2009). Spline smoothing in small area trend estimation and forecasting, *Computational Statistics and Data Analysis*, **53**, 3616–3629.

[33]   Ugarte, M.D.; Militino, A.F. and Goicoa, T. (2008). Prediction error estimators in empirical Bayes disease mapping, *Environmetrics*, **19**, 287–300.

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

[35]   Zhu, L.; Pickle, L.W.; Ghosh, K.; Naishadham, D.; Portier, K.; Chen, H.S.; Kim, H.J; Zou, Z.; Cucinelli, J.; Kohler, B.; Edwards, B.K.; King, J.; Feuer, E.J. and Jemal, A. (2012). Predicting US- and state-level cancer counts for the current calendar year. Part II: Evaluation of spatiotemporal projection methods for incidence, *Cancer*, **118**, 1100–1109.