ENVIRONMETRICS
Environmetrics 2010; **21**: 270–289
Published online 3 July 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.1011

Spatio-temporal modeling of mortality risks using penalized splines

M. D. Ugarte*,†, T. Goicoa and A. F. Militino

Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain

SUMMARY

Analyzing the temporal evolution of the geographical distribution of mortality (or incidence) risks is an important area of research in disease mapping. It might help to better determining the risk factors involved in the studied disease, and to address important epidemiological questions about the stability of the estimated patterns of disease. Traditionally, risk smoothing is carried out using conditional autoregressive (CAR) models but very recently, penalized splines have also been considered in an Empirical Bayes (EB) spatial context to estimate large-scale spatial trends together with region random effects. In this paper, penalized splines for smoothing risks in both the spatial and the temporal dimensions will be applied. The mean squared error (MSE) of the log-risk predictor will be derived allowing for constructing confidence intervals for the risks. To illustrate the procedure mortality data due to brain cancer in continental Spain over the period 1996–2005 are analyzed. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: CAR models; MSE of the log-risk predictor; smoothing risks; brain cancer

1. INTRODUCTION

Mapping mortality or incidence risks is essentially a problem of describing the spatial distribution of risks over a region. Nowadays, the availability of historical high quality mortality registers offers the possibility of going further describing the spatio-temporal distribution of risks. Such distributions highlight the underlying temporal trends and spatial patterns of the disease risks and might be useful for epidemiologists and public health researchers to formulate etiologic hypothesis. The problem of detecting areas with high risks is also crucial to discover possible inequalities in health among regions and then, to implement appropriate health policies concerning allocation of health funds.

The recent literature on space-time disease mapping has been mainly developed under a fully Bayesian framework (see for example, Waller *et al*., 1997; Knorr-Held, 2000; Knorr-Held and Richardson, 2003; Martínez-Beneito *et al*., 2008) by using generally conditional autoregressive (CAR)

*Correspondence to: M. D. Ugarte, Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain.
†E-mail: lola@unavarra.es

Received 4 March 2009
Accepted 7 April 2009

models. There has also been some research on the use of splines in disease mapping from an Empirical Bayes (EB) and a fully Bayes (FB) perspective. MacNab and Dean (2001) consider autoregressive local smoothing across the spatial dimension and B-spline smoothing over the temporal dimension. They estimate the model using the well known penalized quasi-likelihood technique (PQL)(Breslow and Clayton, 1993). However, this model is separable in space and time, a feature not very realistic in practice. Lee and Durbán (2008a) propose the use of penalized splines (P-splines) for smoothing spatial count data. Within the FB setting, MacNab and Gustafson (2007) develop a semiparametric model that facilitates temporal smoothing of risks using regression B-splines. In particular, they illustrate how time trends in small-area relative risks may be explored by using splines which vary in either a spatially structured or spatially unstructured manner. MacNab (2007) also considers Bayesian hierarchical formulations of regression B-splines, smoothing splines, and P-spline methods in disease mapping, but she concludes that they can be heavily influenced by the hyperprior specifications when data are limited. Here, we propose the use of penalized splines with B-spline basis (Eilers and Marx, 1996), and their representation as a mixed model (in fact, as a generalized linear mixed model)(see Currie and Durbán, 2002) for smoothing mortality risks. Smoothing is carried out in three dimensions: longitude, latitude, and time, allowing for a different time evolution of both spatial components (see Lee and Durbán, 2008b). The model is then non-separable and anisotropic as it might provide a different amount of smoothing in each direction. Furthermore, separable and non-separable models with a CAR structure for the spatial and the time effects are also considered for comparison purposes. All models are estimated from an EB approach using PQL.

An additional and important objective of this paper that has been generally ignored in the EB P-spline literature is to provide accurate confidence intervals (bands) for the risks. Ugarte *et al*. (2008) derive second-order correct MSE estimators of the log-risk predictor and build the corresponding confidence intervals for the risks in a spatial context when using CAR models for smoothing risks. Ugarte *et al*. (2009a, 2009b) also analyze Bayesian Decision Rules and EB confidence intervals for detecting areas with high risks in a spatial and spatio-temporal context but again when using CAR models for smoothing risks. In this work, a second-order correct MSE estimator of the log-risk predictor obtained from the P-spline model is derived. It takes account of the variability induced by the estimation of the variance components (smoothing parameters). This allows us to construct confidence intervals for the risks and then, areas with high risks can be detected. The methodology developed in this paper is used to analyze mortality data due to brain cancer in continental Spain over the period 1996–2005.

The rest of the paper is organized as follows. In Section 2, the spatio-temporal P-spline model introduced by Lee and Durbán (2008b) is derived, and a non-separable CAR model is also considered. The PQL technique used to estimate the model parameters is briefly reviewed. In Section 3, the MSE for the log-risk predictor is developed. Section 4 illustrates the techniques with the analysis of brain cancer mortality data from Spain. The paper ends with a discussion.

2. MODELS FOR SPATIO-TEMPORAL COUNT DATA AND THE PQL TECHNIQUE

Traditionally, incidence or mortality data from rare diseases have been modeled using a mixed Poisson model with a conditional autoregressive distribution (CAR) for the region random effects accounting for spatial dependence. One of the possible extensions for analyzing spatio-temporal count data has been the inclusion of a time random effect with a CAR-type distribution (Knorr-Held, 2000). An alternative approach is the use of P-spline models. P-spline models are different to the CAR models commonly used in disease mapping as they smooth global spatio-temporal trends, while traditional CAR models take

DOI: 10.1002/env