



UNIVERSIDAD CARLOS III DE MADRID

working  
papers

Working Paper 08-58  
Statistics and Econometrics Series 20  
November 2008

Departamento de Estadística  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 91 624-98-49

## SMOOTH-CAR MIXED MODELS FOR SPATIAL COUNT DATA

Dae-Jin Lee and María Durbán

### Abstract

---

Penalized splines (P-splines) and individual random effects are used for the analysis of spatial count data. P-splines are represented as mixed models to give a unified approach to the model estimation procedure. First, a model where the spatial variation is modelled by a two-dimensional P-spline at the centroids of the areas or regions is considered. In addition, individual area-effects are incorporated as random effects to account for individual variation among regions. Finally, the model is extended by considering a conditional autoregressive (CAR) structure for the random effects, these are the so called “Smooth-CAR” models, with the aim of separating the large-scale geographical trend, and local spatial correlation. The methodology proposed is applied to the analysis of lip cancer incidence rates in Scotland.

---

**Keywords:** Mixed models; P-splines; Overdispersion; Negative Binomial; PQL; CAR models; Scottish Lip Cancer data.

Universidad Carlos III de Madrid, Department of Statistics, Escuela Politécnica Superior, Campus de Leganés, Madrid, Spain. e-mail addresses: [dae-jin.lee@uc3m.es](mailto:dae-jin.lee@uc3m.es) (Dae-Jin Lee) and [mdurban@est-econ.uc3m.es](mailto:mdurban@est-econ.uc3m.es) (María Durbán). **Acknowledgements:** The research of Dae-Jin Lee and Maria Durbán was supported by Comunidad de Madrid, project CCG06-UC3M/ESP-0856, and Ministerio de Educación y Ciencia, project SEJ2005-06454.

# Smooth-CAR Mixed Models for Spatial Count Data

Dae-Jin Lee<sup>a</sup>, María Durbán<sup>a,\*</sup>

<sup>a</sup>*Universidad Carlos III de Madrid, Department of Statistics, Escuela Politécnica Superior, Campus de Leganés, Madrid, SPAIN*

---

## Abstract

We propose the use of Penalized splines ( $P$ -splines) and individual random effects for the analysis of spatial count data.  $P$ -splines are represented as mixed models to give a unified approach to the model estimation procedure. First, we introduce a model where the spatial variation is modelled by a two-dimensional  $P$ -spline at the centroids of the areas or regions. Additionally, individual area-effects are incorporated as random effects to account for individual variation between regions. Finally, we extend the model by considering a conditional autoregressive (CAR) structure for the random effects, these are the so called “Smooth-CAR” models, with the aim to separate the large-scale geographical trend and the local spatial correlation. We apply the methodology proposed to the analysis of lip cancer incidence rates in Scotland.

*Key words:* Mixed models;  $P$ -splines; Overdispersion; Negative Binomial; PQL; CAR models; *Scottish Lip Cancer* data.

---

---

\* Corresponding author.

*Email addresses:* [dae-jin.lee@uc3m.es](mailto:dae-jin.lee@uc3m.es) (Dae-Jin Lee),  
[mdurban@est-econ.uc3m.es](mailto:mdurban@est-econ.uc3m.es) (María Durbán).

## 1 Introduction

Spatial data are geographically distributed over locations in a map. They represent a continuous phenomena at sampling locations (*geostatistical* data), or over spatial indexed units such as regions or countries where data are aggregated (*lattice* or regional data). In both cases, we are interested in studying the spatial variation on the response across the spatial domain of interest, by modelling the correlation structure. Penalized splines (Eilers and Marx, 1996) are a well established method for smoothing Gaussian and non-Gaussian data in one or more dimensions (Currie et al., 2006). Their representation as mixed models (Currie and Durbán, 2002; Wand, 2003) presents attractive features in spatial statistics: random effects or spatial correlation can be estimated together with a large-scale spatial trend. Kammann and Wand (2003) presents a connection between kriging, and (generalized additive models (Hastie and Tibshirani, 1990) also in the mixed model context.

Other approach to model spatial data are the conditionally autoregressive (CAR) models (Besag, 1974). They have been very popular in the analysis of regional data. These models consider the spatial dependence locally across “neighbouring” areas. Neighbourhoods may be defined in several ways, the most common definition is to consider two or more regions as neighbors if they share a common border; other criteria consider the Euclidean distance between the centroids (see Cressie and Chan, 1989; Besag et al., 1991; Cressie, 1993; Besag and Kooperberg, 1995).

In many applications, it is common to collect count data observed in spatial locations, and assume a Poisson distribution for the counts. In the analysis of spatial count data, hierarchical models (mixed Poisson models) allow the incorporating of area-specific random effects. However, the equality of the mean and variance in the Poisson assumption is, generally, too restrictive when data

present a variance higher than expected. This phenomena is known as overdispersion. When considering spatially indexed count data, this overdispersion phenomena can be viewed as an unobserved heterogeneity caused by an underlying spatial dependency structure, aggregation, or the omission of relevant covariates. Alternatively, Negative Binomial distribution has been used to allow independent modelling of the mean and variance by an additional parameter. Examples of modelling overdispersed data can be found in Lawless (1987); Dean and Lawless (1989); Hinde and Demetrio (1998); Thurston et al. (2000).

In this paper we propose the use of the mixed model approach for spatial counts. In a first model, the spatial variation is modelled by a two-dimensional smooth Penalized spline (or  $P$ -spline) where the centroids of the areas or regions are considered as spatial locations in terms of their geographical longitude and latitude. Then, the counts in each region are assumed have been observed in the centroid of the region. A similar approach would be to consider the interpolation of regional data as continuous surfaces, as it happens in geostatistical methods such as *kriging* or Gaussian Random Fields (Cressie and Chan, 1989; Diggle et al., 1998; Kelsall and Wakefield, 2002). Additionally, individual area-effects are incorporated as random effects to account for individual variation between regions. In a second model, we combine a smooth model and random effects with a CAR structure given by the neighbouring areas. We called these models: “*Smooth-CAR*” models. The aim is to be able to separate the large-scale spatial trend and the small-space regional variation, and estimate both effects simultaneously. We applied the proposed models to the analysis of lip cancer incidence cases in Scotland over the period 1975-1980.

The paper is organized as follows. In sections 2 and 3 we present  $P$ -splines as mixed models with individual random effects as method for the analysis of spatial data. Section 4 extends the previous model by incorporating random

effects with a CAR structure to account for the local spatial correlation. We illustrate and compare these models in Section 5 with the Scottish lip cancer data, and we conclude with a discussion in section 6.

## 2 *P*-spline smoothing for spatial data

### 2.1 *Smoothing with P-splines*

For simplicity, let's suppose normally distributed spatial data  $(\mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{y}_{ij})$ , where  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are respectively geographic longitude and latitude and  $\mathbf{y}$  is the response variable. A smooth model for the data would be given by:

$$\mathbf{y} = f(\mathbf{x}_1, \mathbf{x}_2) + \epsilon = \mathbf{B}\boldsymbol{\theta} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \quad (2.1)$$

where  $\boldsymbol{\theta}$  is a vector of coefficients, and  $\mathbf{B}$  is a regression basis constructed from the covariates  $(\mathbf{x}_1, \mathbf{x}_2)$ . The *P*-spline approach minimizes the penalized sum of squares

$$S(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta} \quad (2.2)$$

where  $\mathbf{P}$  is a discrete penalty matrix which depends on one or more smoothing parameters, and impose smoothness over adjacent coefficients.

When data are non-normal (as in the case of spatial count data), Marx and Eilers (1998) used Penalized splines as an extension of the generalized linear model (GLM) formulation. Assuming that the distribution of  $\mathbf{y}$  belongs to the exponential family, and given a link function  $g(\cdot)$ , then,

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{B}\boldsymbol{\theta}. \quad (2.3)$$

The penalized sum of squares given in equation (2.1) becomes:

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}, \quad (2.4)$$

where  $\ell(\boldsymbol{\theta})$  is the ordinary likelihood. This yields the penalized version of the “scoring algorithm”:

$$(\mathbf{B}'\tilde{\mathbf{W}}_{\delta}\mathbf{B} + \mathbf{P})\hat{\boldsymbol{\theta}} = \mathbf{B}'\tilde{\mathbf{W}}_{\delta}\mathbf{B}\tilde{\boldsymbol{\theta}} + \mathbf{B}'(\mathbf{y} - \tilde{\boldsymbol{\mu}}), \quad (2.5)$$

where  $\mathbf{W}_{\delta}$  is a diagonal matrix, with diagonal elements  $w_{ii}^{-1} = (\partial\eta_i/\partial\mu_i)^2\text{var}(y_i)$ ,  $\tilde{\boldsymbol{\theta}}$  represents an approximate solution, and  $\hat{\boldsymbol{\theta}}$  represents the improved approximation.

### 2.1.1 Basis and Penalties for spatial data

As we saw above,  $P$ -spline regression depends on a regression basis and a penalty matrix which controls the smoothness of the fit. The literature reflects the variety of approaches to the construction of such basis and penalties (see for example: Eilers et al. (2006); Wand (2003); Lang and Brezger (2004)). Each of them have advantages and disadvantages, and perform differently depending on the data analyzed and context where they are used (see Welham et al. (2007) for details). In the case of two-dimensional smoothing (as it is the case of spatial data) the election of basis and penalty is even more important and the differences between the approaches are significant. Ruppert et al. (2003) proposed the use radial basis functions. This basis has the limitation of being an isotropic smoother and also, the selection of knots to construct the basis is not trivial. Other authors such as Lang and Brezger (2004); Wood (2006); Currie et al. (2006) propose the use of tensor product of  $B$ -spline basis with equally spaced knots. This is the approach we take in this article.

In the univariate case, given a single covariate  $\boldsymbol{x}' = (x_1, \dots, x_n)$ , a  $B$ -spline basis  $\mathbf{B} = (B_1(\boldsymbol{x}), B_2(\boldsymbol{x}), \dots, B_k(\boldsymbol{x}))$  is an  $n \times k$  matrix of  $B$ -splines ( $k$  depends on the number of knots and the degree of the  $B$ -spline). The extension to the two-dimensional case depends on the structure of the data. If we have scattered spatial data (see Eilers et al. (2006) for details), the basis is constructed

Table 2.1

Tensor products of  $B$ -splines basis for array or scattered data.

<b>Data</b>	<b>regular GRID/array</b>	<b>scattered data</b>
<b>e.g.</b>	(image data)	(geostatistical data)
<b>Product</b>	Kronecker	row-wise Kron.
<b>Basis</b>	$\mathbf{B}_2 \otimes \mathbf{B}_1$	$\mathbf{B}_2 \square \mathbf{B}_1$

from the tensor product of marginal  $B$ -spline basis defined as the Box-Product or “row-wise” Kronecker product of the individual basis, denoted by  $\square$ :

$$\mathbf{B} = \mathbf{B}_2 \square \mathbf{B}_1 = (\mathbf{B}_2 \otimes \mathbf{1}_{c_1}) \odot (\mathbf{1}_{c_2} \otimes \mathbf{B}_1) , \quad (2.6)$$

where  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are the  $B$ -spline basis along the longitude ( $\mathbf{x}_1$ ) and latitude ( $\mathbf{x}_2$ ) coordinates of dimensions  $n \times c_1$  and  $n \times c_2$ . The basis  $\mathbf{B}$  is of dimension  $n \times c_1 c_2$ , and the operator  $\odot$  is the Hadamard or “element-wise” matrix product and  $\mathbf{1}_{c_1}$  and  $\mathbf{1}_{c_2}$  are vectors of ones of length  $c_1$  and  $c_2$ .

When data are in a regular grid, the basis is defined as the Kronecker product of the marginal basis. Although  $P$ -splines are low rank smoothers, when the data set is large, fitting this type of models could be computationally intensive. This might be one of the reasons why  $P$ -splines, and splines in general, haven’t been a popular method to model spatial data. However, recently, Currie et al. (2006) developed algorithms that reduce considerably the computational time. These methods allow the use of these type of smoothers when the amount of data is large (which is often the case of spatial and spatio-temporal data). A summary of the two-dimensional basis can be found in Table 2.1.

The other aspect in which  $P$ -splines for two-dimensional data differ is the penalty applied to the coefficients. Currie et al. (2004, 2006) among others, proposed a penalty matrix based on the penalties associated with each of the marginal basis, i.e., if  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are the penalties corresponding to the basis

in each dimension, then,

$$\mathbf{P} = \lambda_2 \mathbf{P}_2 \otimes \mathbf{I}_{c_1} + \lambda_1 \mathbf{I}_{c_2} \otimes \mathbf{P}_1 , \quad (2.7)$$

where  $\lambda_1$  and  $\lambda_2$  are smoothing parameters which tune the smoothness in each direction, separately. Therefore, this penalty allow for anisotropy. This is a very important point when covariates are measured in different scales (such as space and time), or when different degree of smoothing is needed in each dimension. Different authors use different definitions of the individual penalties  $\mathbf{P}_i$  ( $i = 1, 2$ ), Wood (2006) uses the usual spline integrated squared second derivative penalty, while Currie et al. (2006) use a discrete penalty based on difference matrices  $\mathbf{D}_i$  to impose smoothness over adjacent coefficients,  $\mathbf{P}_i = \mathbf{D}'_i \mathbf{D}_i$ , in the case of differences of order two, we have

$$\mathbf{D}_i = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix} \quad (2.8)$$

Therefore the penalty is defined by

$$\mathbf{P} = \lambda_2 \mathbf{D}'_2 \mathbf{D}_2 \otimes \mathbf{I}_{c_1} + \lambda_1 \mathbf{I}_{c_2} \otimes \mathbf{D}'_1 \mathbf{D}_1 . \quad (2.9)$$

The Bayesian analogous of  $P$ -splines consider the difference matrix  $\mathbf{D}$  in (2.8) as random walks of order  $m$  used as priors for the regression coefficients (Lang and Brezger, 2004). In the spatial context, they specify Markov Random Field priors for the  $B$ -spline regression coefficients  $\boldsymbol{\theta}$  (based on Besag and Kooperberg (1995)) constructed from the four nearest neighbors. They also suggest a prior based on the Kronecker product penalty:

$$\mathbf{D}'_1 \mathbf{D}_1 \otimes \mathbf{D}'_2 \mathbf{D}_2 \quad (2.10)$$

In both cases the model is isotropic and has a rank deficiency problem. Therefore, we will take the approach given in Currie et al. (2006) and use the penalty



defined in (2.9)

## 2.2 Mixed Models representation

The connection between non-parametric regression and mixed models goes back to the early 90s (Speed, 1991). Authors like Verbyla et al. (1999) and Brumback and Rice (1998) developed the topic of smoothing with mixed models in the context of smoothing splines, and Wand (2003) in the context of  $P$ -splines with truncated lines as basis. However, the mixed model representation of  $P$ -splines as mixed models using  $B$ -splines as basis was not studied until recently (Currie and Durbán, 2002). The mixed model approach of  $P$ -splines allow us to relax the linearity assumptions commonly used in the mixed model context and combine, for example, a two-dimensional smooth surface with individual random effects with a correlation structure.

The aim is to set a new basis which allows the representation of Eq. (2.1) and its associated penalty (2.9) into a mixed model such as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{G}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\sigma}^2\mathbf{I}), \quad (2.11)$$

where  $\mathbf{G}$  is a variance component matrix depending on one or more smoothing parameters  $\boldsymbol{\lambda}$ , and  $\mathbf{X}$  and  $\mathbf{Z}$  are respectively the fixed and random effects matrices.

Following a similar approach to Currie et al. (2006), the idea is to use the singular value decomposition (SVD) of  $\mathbf{P}$  in Eq. (2.9) to partition the difference penalty ( $\mathbf{D}'_1\mathbf{D}_1$  of dimension  $c_1 \times c_1$ ) into a null penalty (for the fixed part) and a diagonal penalty (for the random part). Then,  $\mathbf{D}'_1\mathbf{D}_1$  is decomposed as  $\mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}'_1$ , where  $\boldsymbol{\Sigma}_1$  is a diagonal matrix, whose diagonal elements are the eigenvalues of  $\mathbf{D}'_1\mathbf{D}_1$ .

We can partition the matrix  $\mathbf{U}_1 = [\mathbf{U}_{1s} : \mathbf{U}_{1n}]$ , where  $\mathbf{U}_{1s}$  corresponds to the

non-zero eigenvalues and  $\mathbf{U}_{1n}$  to the zero eigenvalues. Considering a second order penalty,  $\mathbf{\Sigma}_1$  has two zero eigenvalues and  $\mathbf{U}_{1n}$  has two columns. Then, we denote  $\tilde{\mathbf{\Sigma}}_1$  as the diagonal  $(c_2-2) \times (c_2-2)$  matrix of positive eigenvalues of  $\mathbf{\Sigma}_1$ . We have similar definitions for  $\mathbf{U}_{2s}$ ,  $\mathbf{U}_{2n}$  and  $\tilde{\mathbf{\Sigma}}_2$  from the SVD decomposition of  $\mathbf{D}'_2\mathbf{D}_2$ .

We need to find a transformation  $\mathbf{T}$ , such that  $\mathbf{BT} = [\mathbf{X} : \mathbf{Z}]$ , where  $\mathbf{X}$  and  $\mathbf{Z}$  are orthogonal. This matrix  $\mathbf{T}$  is not unique, in fact any *one-to-one* transformation such that the model is reparameterized as  $\mathbf{B}\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ , can be applied and the coefficients as  $\boldsymbol{\theta} = \mathbf{T}(\boldsymbol{\beta} : \boldsymbol{\alpha})'$ .

We define  $\mathbf{T}$  as a orthogonal transformation matrix given by

$$\mathbf{T} = (\mathbf{U}_{2n} \otimes \mathbf{U}_{1n} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1n} : \mathbf{U}_{2n} \otimes \mathbf{U}_{1s} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1s}), \quad (2.12)$$

with this transformation, the penalty term  $\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta} = \boldsymbol{\omega}'\mathbf{T}'\mathbf{P}\mathbf{T}\boldsymbol{\omega}$ , where  $\boldsymbol{\omega}' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$  and the coefficients  $\boldsymbol{\beta} = (\mathbf{U}_{2n} \otimes \mathbf{U}_{1n})'\boldsymbol{\theta}$  and  $\boldsymbol{\alpha} = (\mathbf{U}_{2s} \otimes \mathbf{U}_{1s})'\boldsymbol{\theta}$ . Now,  $\mathbf{P}(\mathbf{U}_{2n} \otimes \mathbf{U}_{1n}) = 0$ , so there is no penalty on the fixed part and the penalty becomes  $\boldsymbol{\alpha}'\mathbf{F}\boldsymbol{\alpha}$  for some  $\mathbf{F}$ .

It is straightforward to show that the penalty  $\mathbf{F}$  in the mixed model is the block diagonal matrix:

$$\mathbf{F} = \begin{pmatrix} \lambda_2 \tilde{\mathbf{\Sigma}}_2 \otimes \mathbf{I}_2 & & \\ & \lambda_1 \mathbf{I}_2 \otimes \tilde{\mathbf{\Sigma}}_1 & \\ & & \lambda_1 \mathbf{I}_{c_2-2} \otimes \tilde{\mathbf{\Sigma}}_1 + \lambda_2 \tilde{\mathbf{\Sigma}}_2 \otimes \mathbf{I}_{c_1-2} \end{pmatrix} \quad (2.13)$$

where  $\mathbf{F} = \mathbf{T}\mathbf{P}\mathbf{T}'$  and the covariance matrix of the random effects is  $\mathbf{G} = \sigma^2 \mathbf{F}^{-1}$ .

Using the properties of “row-wise” Kronecker product (Liu, 1999) it is easy to show that for scattered data we may take define the matrix of fixed effects as

$$\mathbf{X} = \mathbf{X}_2 \square \mathbf{X}_1, \quad (2.14)$$

where  $\mathbf{X}_2 = [\mathbf{1}_n : \mathbf{x}_2]$ ,  $\mathbf{X}_1 = [\mathbf{1}_n : \mathbf{x}_1]$ , and  $\mathbf{1}_n$  is a vector of ones of length  $n$ . And the random effects matrix as:

$$\mathbf{Z} = [\mathbf{Z}_2 \square \mathbf{X}_1 : \mathbf{X}_2 \square \mathbf{Z}_1 : \mathbf{Z}_2 \square \mathbf{Z}_1] , \quad (2.15)$$

where  $\mathbf{Z}_2 = \mathbf{B}_2 \mathbf{U}_{2s}$  and  $\mathbf{Z}_1 = \mathbf{B}_1 \mathbf{U}_{1s}$ .

We can expand  $\mathbf{X}$  and  $\mathbf{Z}$  in Equations (2.14) and (2.15) as

$$\mathbf{X} = (\mathbf{1}_n : \mathbf{1}_n \square \mathbf{x}_1 : \mathbf{x}_2 \square \mathbf{1}_n : \mathbf{x}_2 \square \mathbf{x}_1) \quad (2.16)$$

$$\mathbf{Z} = (\mathbf{Z}_2 \square \mathbf{1}_n : \mathbf{Z}_2 \square \mathbf{x}_1 : \mathbf{1}_n \square \mathbf{Z}_1 : \mathbf{x}_2 \square \mathbf{Z}_1 : \mathbf{Z}_2 \square \mathbf{Z}_1) \quad (2.17)$$

This partition allows the representation of the fitted surface in terms of a sum of three components: one for  $\mathbf{x}_1$  (second block of  $\mathbf{X}$  in (2.16) and first block of  $\mathbf{Z}$  in (2.17)), a component for  $\mathbf{x}_2$  (third block of  $\mathbf{X}$  in (2.16) and third block of  $\mathbf{Z}$  in (2.17)) and an interaction component which depends on both covariates and given by the four remaining terms.

Given the new basis (2.16), (2.17) and the new penalty (2.13), the penalized sum of squares (2.2) transforms in the mixed model into

$$S(\boldsymbol{\beta}, \boldsymbol{\alpha}; \mathbf{y}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) + \boldsymbol{\alpha}'\mathbf{F}\boldsymbol{\alpha} \quad (2.18)$$

In the context of the mixed models, smoothing parameters  $\boldsymbol{\lambda}_1$  and  $\boldsymbol{\lambda}_2$  may be selected by maximizing the residual log-likelihood (REML) of Patterson and Thompson (1971), i.e.  $\ell(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \sigma^2)$ :

$$-\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2} \mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y} \quad (2.19)$$

where  $\mathbf{V} = \sigma^2 \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$ . It may be shown that

$$|\mathbf{V}| = \sigma^{2n} |\mathbf{G}| |\mathbf{G}^{-1} + \frac{1}{\sigma^2} \mathbf{Z}'\mathbf{Z}| \quad (2.20)$$

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2} \left( \mathbf{I} - \mathbf{Z}(\sigma^2 \mathbf{G}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \right) \quad (2.21)$$

Equations (2.20) and (2.21) reduce considerably the computational time. We

do not need to compute  $\mathbf{V}$  (of dimension  $n \times n$ ) and its inverse, but a matrix of dimension  $c_1 c_2 \times c_1 c_2$  matrix in both cases.

The estimates of the coefficients  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}$  follow from standard mixed model theory

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (2.22)$$

$$\hat{\boldsymbol{\alpha}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.23)$$

This methodology can be applied to non-Gaussian data in the Generalized Linear Mixed Model (GLMM) context. This approach consist of two stages: (i) reparameterizing the linear predictor and (ii) estimating the model parameters, which will be done using the penalized quasi-likelihood (PQL) approach of Breslow and Clayton (1993).

### 3 Modelling Overdispersion with $P$ -splines

Count data often present extra Poisson variation caused by an unobserved heterogeneity, this phenomenon is known as *overdispersion*, and implies that the variance of the data is greater than its mean, sometimes much larger. If the data are overdispersed but this is ignored, we are overweighting the data and consequently underestimating their variability. The problem of overdispersion has been studied by many authors from several points of view (see for example, Lawless (1987); Dean and Lawless (1989); Hinde and Demetrio (1998); Thurston et al. (2000)). It is natural to assume that the number of occurrence of events are realizations of a Poisson distribution and can be fitted by a Poisson regression model. Currie et al. (2004) showed the penalized GLM approach for Poisson data in two dimensions. We show present here how  $P$ -splines can be used to model spatial count data in the presence of overdispersion. Spatial variation is modelled by a smooth surface, and we take

two approaches to account for overdispersion, which are detailed in next two subsections: (i) introduce individual random effects and (ii) use of negative binomial distribution. Using the mixed model representation of  $P$ -splines we can estimate efficiently the spatial effects and overdispersion.

### 3.1 The PRIDE approach

Perperoglou and Eilers (2007) gave an approach based on penalized likelihood, using individual random effects which adds extra parameters ( $\boldsymbol{\gamma}$ ) to the linear predictor of a Poisson GLM (with log link) for each observation.

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\gamma}\mathbf{I}, \quad \boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \kappa^{-1}\mathbf{I}) \quad (3.1)$$

This model is called PRIDE (“*Penalized Random Individual Dispersion Effects*”). As showed in Section 2.2, model (3.1) can be reparameterized as a mixed model so the linear predictor becomes

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\gamma}\mathbf{I}, \quad \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \kappa^{-1}\mathbf{I}), \quad (3.2)$$

where  $\boldsymbol{\gamma}$  is a  $n \times 1$  random effect vector which provides a device to absorb the overdispersion which causes the extra variability. Given that the model has more parameters than observations we add a ridge penalty on  $\boldsymbol{\gamma}$  to maintain the identifiability of the model which shrinks  $\boldsymbol{\gamma}$  to zero.

Using PQL we obtain the following set of equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{G}^{-1} + \mathbf{Z}'\mathbf{W}\mathbf{Z} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}\mathbf{X} & \mathbf{W}\mathbf{Z} & \kappa\mathbf{I} + \mathbf{W} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{z} \\ \mathbf{Z}'\mathbf{W}\mathbf{z} \\ \mathbf{W}\mathbf{z} \end{bmatrix}, \quad (3.3)$$

where  $\mathbf{z}$  is the *working vector*,  $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1}(y - \boldsymbol{\mu})$ , and  $\mathbf{W}$  is the diagonal matrix of weights,  $\mathbf{W} = \text{diag}(\boldsymbol{\mu})$  and  $\boldsymbol{\mu} = \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\gamma}\mathbf{I})$ . Equation (3.3) gives a very large system of equations but it is possible to reduce it by

eliminating  $\gamma$ , by

$$\gamma = \frac{\mathbf{W}}{\mathbf{W} + \kappa \mathbf{I}} (z - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) \quad (3.4)$$

And defining a modified weight matrix:

$$\mathbf{W}^* = \frac{\kappa \mathbf{W}}{\mathbf{W} + \kappa \mathbf{I}}. \quad (3.5)$$

Given (3.5) it is easy to show that (3.3) becomes:

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}^*\mathbf{X} & \mathbf{X}'\mathbf{W}^*\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}^*\mathbf{X} & \mathbf{G}^{-1} + \mathbf{Z}'\mathbf{W}^*\mathbf{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}^*z \\ \mathbf{Z}'\mathbf{W}^*z \end{bmatrix} \quad (3.6)$$

These equations yield the estimates in a weighted linear Gaussian model with error variance  $\sigma^2 = 1$ ,  $\hat{\boldsymbol{\beta}}$  and  $\hat{\boldsymbol{\alpha}}$  are estimated as in (2.22), (2.23) and  $\hat{\gamma}$  as in (3.4). Then, conditional on the estimates of the coefficients, we estimate  $\boldsymbol{\lambda}_1$ ,  $\boldsymbol{\lambda}_2$  and  $\kappa$  by REML as in Equation (2.19).

Now the variance components matrix  $\mathbf{V}$  takes the form

$$\mathbf{V} = \mathbf{W}^{*-1} + \mathbf{Z}\mathbf{G}\mathbf{Z}' \quad (3.7)$$

Then, again, conditional on the estimates obtained above, we use REML to estimate the variance components and iterate (2.23), (2.22) and (2.19) till convergence.

### 3.2 Negative Binomial regression

The negative binomial regression is used to estimate count models in the presence of overdispersion. The negative binomial distribution permits more flexible modelling of the variance than the Poisson case where the assumption of equal mean and variance is inappropriate. The negative binomial distribution is derived by letting the mean of the Poisson distribution vary according to a fixed parameter  $\zeta_i$  given by the Gamma distribution.

Let  $\mathbf{y}_i$  be a set of count data and

$$\begin{aligned}\mathbf{y}_i | \zeta_i &\sim \mathcal{Poisson}(\zeta_i \mu_i), \\ \zeta_i &\sim \frac{1}{\kappa} \mathcal{Gamma}(\kappa),\end{aligned}$$

the marginal distribution of  $\mathbf{y}_i$  is then the negative binomial with mean  $\mu_i$  and variance  $\mu_i + \mu_i^2/\kappa$ , where  $\kappa$  is a dispersion or shape parameter.

$$\mathbf{y}_i \sim \text{Neg Bin}(\mu_i, \kappa)$$

The estimation of this model includes a parameter  $\kappa$  which is an estimate of the degree of overdispersion. However, the fact that the negative binomial has two parameters and it is not in the exponential family, makes it more difficult to extend the methodology developed for Poisson data.

The negative binomial distribution is given by

$$P(\mathbf{Y}_i = \mathbf{y}_i | \mu_i, \kappa) = \binom{\mathbf{y}_i + \kappa - 1}{\mathbf{y}_i} \left( \frac{\mu_i}{\kappa + \mu_i} \right)^{\mathbf{y}_i} \left( \frac{\kappa}{\kappa + \mu_i} \right)^\kappa,$$

where  $\binom{\mathbf{y}_i + \kappa - 1}{\mathbf{y}_i} = \frac{\Gamma(\mathbf{y}_i + \kappa)}{\Gamma(\kappa) \Gamma(\mathbf{y}_i + 1)}$  and  $\mu_i = E[\mathbf{Y}_i]$ .

Under this model the variance of  $\mathbf{Y}_i$  is  $\mu_i + \mu_i^2/\kappa$ . Note that, for large values of  $\kappa$ , this model reduces to the Poisson case.

It is possible to write the log-likelihood as an exponential family:

$$\begin{aligned}\ell(\mu_i, \kappa | \mathbf{Y}_i) &= \mathbf{Y}_i \ln \left( \frac{\mu_i}{\mu_i + \kappa} \right) + \kappa \ln \left( \frac{\kappa}{\mu_i + \kappa} \right) \\ &\quad + \ln \Gamma(\mathbf{Y}_i + \kappa) - \ln \Gamma(\kappa) - \ln \Gamma(\mathbf{Y}_i + 1) + \kappa \ln \kappa.\end{aligned}\tag{3.8}$$

Thus, in the case of  $P$ -splines for a GLM, the linear predictor is  $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\theta}$  and in the mixed model context, where  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ , the joint density for a Negative Binomial GLMM, with log link is

$$\begin{aligned}f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \kappa) &= \exp \left[ \mathbf{y}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}) - \log \{ \kappa \mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}) \} \right] - \kappa \mathbf{1}' \log \{ \kappa \mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}) \} \\ &\quad + \exp \left[ n \kappa \log(\kappa) + \mathbf{1}' \log(\mathbf{y} + \kappa \mathbf{1}) - n \log(\Gamma(\kappa)) \right],\end{aligned}\tag{3.9}$$

Table 3.1  
Comparison of count data regression models.

Model	log link	Inverse link	weight matrix
Poisson	$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$	$\boldsymbol{\mu} = e^{\boldsymbol{\eta}}$	$\mathbf{W} = \text{diag}(\boldsymbol{\mu})$
PRIDE	$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\gamma}\mathbf{I}$	$\boldsymbol{\mu} = e^{\boldsymbol{\eta}}$	$\mathbf{W}^* = \frac{\kappa \text{diag}(\boldsymbol{\mu})}{\text{diag}(\boldsymbol{\mu}) + \kappa\mathbf{I}}$
Neg. Binomial	$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$	$\boldsymbol{\mu} = e^{\boldsymbol{\eta}}$	$\mathbf{W} = \kappa \text{diag}\left(\frac{\boldsymbol{\mu}}{\kappa + \boldsymbol{\mu}}\right)$

and the penalized likelihood:

$$\begin{aligned}
\ell_p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \kappa, \sigma_{\boldsymbol{\alpha}}^2) = & \mathbf{y}' [\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} - \log \{\kappa\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha})\}] \\
& - \kappa\mathbf{1}' \log \{\kappa\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha})\} + n \kappa \log(\kappa) + \mathbf{1}' \log \Gamma(\mathbf{y} + \kappa\mathbf{1}) \\
& - n \log \Gamma(\kappa) - \frac{1}{2\sigma_{\boldsymbol{\alpha}}^2} \boldsymbol{\alpha}' \mathbf{F} \boldsymbol{\alpha}.
\end{aligned} \tag{3.10}$$

It can be shown that the estimation of fixed and random effects and variance components in negative binomial GLMM is done as in the case of Poisson data but the matrix of weights is given by

$$\mathbf{W} = \kappa \text{diag} \left( \frac{\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha})}{\kappa\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha})} \right). \tag{3.11}$$

Table 3.1 shows the main differences in count data regression models, note that the weights are very similar to the ones in the previous Section 3.1 (PRIDE approach), but in this case, we can estimate the individual random effect  $\boldsymbol{\gamma}$ , which also appears in  $\mathbf{W}^*$  through the linear predictor  $\boldsymbol{\eta}$ .

#### 4 Smooth models with CAR structure

The most popular approach in modelling spatial dependency structure for lattice or regional data are the conditionally autoregressive (CAR) models introduced by Besag (1974). These hierarchical models allow both spatially structured variability and unstructured heterogeneity by assuming a prior distribution for the spatial effects considering the neighboring regions. These



models have been widely used in the context of regional data (Besag and Green (1993); Leroux et al. (1999); Dean et al. (2001); Congdon (2006); Wakefield (2007) among others).

In this section we propose to use simultaneously the  $P$ -spline approach given in previous sections, and a conditionally autoregressive (CAR) structure. We call these models “*Smooth-CAR*” models. The smooth component let us model the spatial trend along larger geographical distances, and the local (non-smooth) correlation is taken into account by means of a CAR component, so we could separate the global trend and the purely individual regional effect. The mixed model representation of  $P$ -splines allows us to fit the model in a GLMM. This is a very challenging task, since it is not clear whether both effects are identifiable and further research still needs to be done, but as we will see in the next section, in the example analyzed, this method performs better than the traditional spatial models and it gives a clearer picture of the spatial variation in the data.

The model is

$$\boldsymbol{\eta} = \underbrace{\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}}_{f(x_1, x_2)} + \underbrace{\mathbf{b}}_{CAR}, \quad (4.1)$$

where  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$  corresponds to the mixed model representation of the  $2d$   $P$ -spline, and now the regional random effect  $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}_b)$ , has covariance matrix given by a CAR model.

The basic spatial (intrinsic) CAR model (Besag et al., 1991) uses the adjacencies to define neighbourhoods in a conditional specification of the model. This model considers the covariance matrix  $\mathbf{G}_b$  as a sum of two separate variance components to represent both spatial and non-spatial correlation. One component to model the spatially-structured variation, and another to model the unstructured or individual region-level heterogeneity in the data. In this case,

$\mathbf{G}_b$  has the form

$$\mathbf{G}_b = \sigma_s^2 \mathbf{Q}^- + \kappa^{-1} \mathbf{I} \quad (4.2)$$

where  $\mathbf{Q} = \{q_{i,j}\}$  is a  $n \times n$  matrix determined by the neighborhood structure of the regions. And  $\mathbf{Q}^-$  denotes the Moore-Penrose Generalized inverse of matrix  $\mathbf{Q}$ . The  $i^{th}$  diagonal elements of  $\mathbf{Q}$  are the number of neighbors in the  $i^{th}$  region. The elements out of the diagonal are

$$q_{i,j} = \begin{cases} -1 & \text{if } i^{th} \text{ and } j^{th} \text{ regions are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

Alternative formulations of Besag's model (4.2) have been proposed in the literature. For example, Leroux et al. (1999) adopts a prior specification of the random effects  $\mathbf{b}$  with covariance matrix  $\mathbf{G}_b$  given by:

$$\mathbf{G}_b = \sigma_s^2 (\phi \mathbf{Q} + (1 - \phi) \mathbf{I})^{-1} , \quad (4.3)$$

where  $\sigma_s^2$  is the random effects variance and  $\phi$  is a spatial autocorrelation parameter. Dean et al. (2001) assumes a covariance matrix for the random effects:

$$\mathbf{G}_b = \sigma_s^2 (\phi \mathbf{Q}^- + (1 - \phi) \mathbf{I}) . \quad (4.4)$$

Note that (4.4) is a reparameterization of the model (4.2). In both models (4.3) and (4.4) the covariance parameters are identifiable and  $\phi$ , measures the relative weight between *structured* and *unstructured* variability,  $0 \leq \phi \leq 1$ .

When  $\phi = 1$ , all the overdispersion is due to the spatial correlation so there is no unstructured heterogeneity and model (4.4) is equivalent to the intrinsic CAR model in (4.2). When  $\phi = 0$ , there is an absence of spatial correlation in the data and the overdispersion is not caused by a spatial heterogeneity. If  $0 < \phi < 1$ , the random effects are correlated and the data presents a combination of spatial structured and unstructured component.

The smooth-CAR model (4.1), can be reformulated as

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{u}, \quad \text{with } \mathbf{Z}^* = [\mathbf{Z} : \mathbf{I}] \quad (4.5)$$

and the random effect  $\mathbf{u} = (\boldsymbol{\alpha} : \mathbf{b})'$  has covariance matrix:

$$\mathbf{G}_u = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_b \end{pmatrix}. \quad (4.6)$$

which contains the smoothing parameters ( $\boldsymbol{\lambda}_1$  and  $\boldsymbol{\lambda}_2$ ) and CAR ( $\kappa$  and  $\boldsymbol{\sigma}_b^2$ ) variance components. The estimation of the model parameters can also be done by PQL and REML. The matrix  $\mathbf{V}$  becomes now  $\mathbf{W}^{-1} + \mathbf{Z}^*\mathbf{G}_u\mathbf{Z}^{*'}$  and the random effects  $\mathbf{u}$  is estimated as

$$\hat{\mathbf{u}} = \mathbf{G}_u\mathbf{Z}^{*'}\mathbf{V}^{-1}(\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (4.7)$$

## 5 Application to Scottish Lip Cancer data

The data set consists on the observed ( $\mathbf{y}$ ) and expected ( $\mathbf{e}$ ) number of cases of lip cancer registered in 56 counties in Scotland during the period 1975-1980. This data set has been analyzed several times in the literature (see Wakefield (2007) for a detailed review). Clayton and Kaldor (1987) analyzed the observed and expected counts using Empirical Bayes estimation, and used several alternatives for the distribution of the random effects. Breslow and Clayton (1993) proposed a conditional independent Poisson model, where the random effect is modelled by Gaussian intrinsic autoregression. A different approach is taken by Yasui and Lele (1997), the hierarchical modelling for spatial disease rates is based on estimating functions. This method led to simpler computations as in the  $P$ -splines case, both approaches are very attractive when data sets are large. The models presented by Dean et al. (2001) and Militino et al. (2001) are a reparametrization of Besag (1984), and allow for the determination of the relative weights between spatial and unstructured variation (these models

have already been presented in the previous section). Finally, in the last few years, Congdon (2006) used a generalized additive form that allows regression to vary over regions, and Congdon (2007) considered continuous and discrete priors that account for risks that are discordant with those of neighbouring areas.

We fitted several models to this data set:

(i) Smooth  $P$ -spline models

\*  $\boldsymbol{\eta} = \log(\mathbf{e}) + f(\text{lon}, \text{lat})$ , where  $\log(\mathbf{e})$  is the offset term (**Poisson** model)

\*  $\boldsymbol{\eta} = \log(\mathbf{e}) + f(\text{lon}, \text{lat}) + \boldsymbol{\gamma}\mathbf{I}$ ,  $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \kappa^{-1}\mathbf{I})$ , (**PRIDE** model), and

\* The **Negative Binomial** version of the Poisson model presented in section 3.2.

(ii) Hierarchical CAR models:  $\boldsymbol{\eta} = \log(\mathbf{e}) + \mathbf{X}\boldsymbol{\beta} + \mathbf{b}$ ,  $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}_b)$ , with

$$\mathbf{G}_b = \sigma_s^2 \mathbf{Q}^- + \kappa^{-1} \mathbf{I} \quad (\mathbf{Besag} \text{ model})$$

$$\mathbf{G}_b = \sigma_s^2 (\phi \mathbf{Q}^- + (1 - \phi) \mathbf{I}) \quad (\mathbf{Dean} \text{ model})$$

$$\mathbf{G}_b = \sigma_s^2 (\phi \mathbf{Q} + (1 - \phi) \mathbf{I})^{-1} \quad (\mathbf{Leroux} \text{ model})$$

In order to compare the proposed models we use the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC):

$$\text{AIC} = \text{Dev} + 2 \times \text{df} \quad (5.1)$$

$$\text{BIC} = 2 \times \text{Dev} + \log(n) \times \text{df} \quad (5.2)$$

where  $\mathbf{df}$  are the effective dimension of the model. This is a measure of the complexity of the fitted model, calculated as the trace of the so-called Hat matrix,  $\text{tr}(\mathbf{H})$  as described in (Hastie and Tibshirani, 1990), and  $\mathbf{H}$  is defined such that  $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ .

Both AIC and BIC correct the fit of the model for the number of parameters involved in the model's estimation. **Dev** is the *deviance* defined to be twice

Table 5.1  
Comparisons of fitted models to Scottish Lip Cancer data.

Model		Parameters					AIC	BIC	df
		$\lambda_1$	$\lambda_2$	$\sigma_s^2$	$\kappa^{-1}$	$\phi$			
<b>Smooth:</b>	<i>Poisson</i>	11.75	3.63	-	-	-	114.04	228.46	15.90
	<i>PRIDE</i>	30.12	5.50	-	0.12	-	89.64	180.08	31.73
	<i>PRIDE*</i>	10.45		-	0.12	-	89.82	180.46	32.31
	<i>Neg. Bin</i>	8.45	1.34	-	0.10	-	72.63	145.56	11.91
<b>CAR:</b>	<i>Besag</i>	-	-	0.78	$10^{-6}$	-	89.36	179.56	32.78
	<i>Dean</i>	-	-	0.78	-	0.99	89.36	179.56	32.78
	<i>Leroux</i>	-	-	0.78	-	0.99	89.36	179.56	32.78
<b>Smooth-CAR:</b>	<i>Besag</i>	30.40	18.28	0.55	-	-	87.48	175.75	30.67
	<i>Dean</i>	30.37	18.21	0.55	-	0.99	87.49	175.77	30.67
	<i>Leroux</i>	30.11	16.37	0.53	-	0.97	87.46	175.70	30.64

the difference between of the maximum log-likelihood and the likelihood of the fitted model and is a measure of the discrepancy of the fitted and actual data. For the CAR models, we follow an Empirical instead of fully Bayes approach (Clayton and Kaldor, 1987) in order to ease model comparisons. It should be noticed that in the case of the Negative Binomial model, the adequacy of the fitted model must be carefully considered since we are assuming different distribution for the data. The definition of the deviance in this case is different as in Poisson distribution (for details see Cameron and Trivedi, 1998).

The results obtained are summarized in Table 5.1. The smooth surface in the **Poisson** model is fitted using two-dimensional  $P$ -splines, where the  $B$ -splines basis was constructed from marginal basis, the number of knots was 15 for each basis and the penalties had order two. The PRIDE model incorporates the spatial random effects ( $\gamma$ ) for each of the 56 counties which allowed us to consider individual characteristics of each county and the possible unstructured variation. The estimation of the spatial effects resulted in the higher values of the

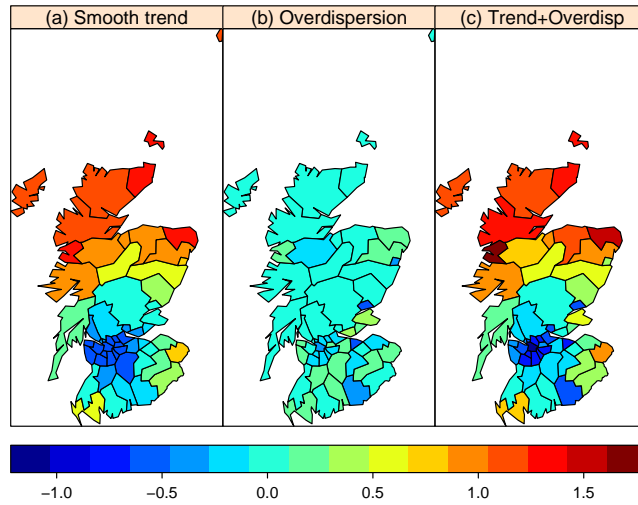


Fig. 5.1. PRIDE Model: (a) Spatial Smooth Trend ( $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ ), (b) Overdispersion individual random effects ( $\boldsymbol{\gamma}$ ) and (c) the sum of trend and overdispersion effects.

effective dimension of the PRIDE model respect to Negative Binomial and Poisson models. Figure 5.1 illustrates the smooth large-scale spatial trend of PRIDE model and the unstructured variation between counties. It can be seen clearly an increasing trend from the more central counties to the ones on the coast, and also from south to north. We have also fitted an isotropic version (both smoothing parameters are equal) of the PRIDE model (See PRIDE\* in Table 5.1), the AIC criteria is slightly lower for the anisotropic model, although there is not much difference. However, it is possible that, even in the situation where both covariates (longitude and latitude) are measured in the same scale, using a single smoothing parameter might not be the appropriate choice.

*Neighbourhoods* are defined by several criteria, depending on the shape of the lattice, for example, the distance between the centroids of the regions, bordering regions or sharing a common border with a given region, (see Cressie and Chan, 1989; Besag et al., 1991; Besag and Kooperberg, 1995). However, when

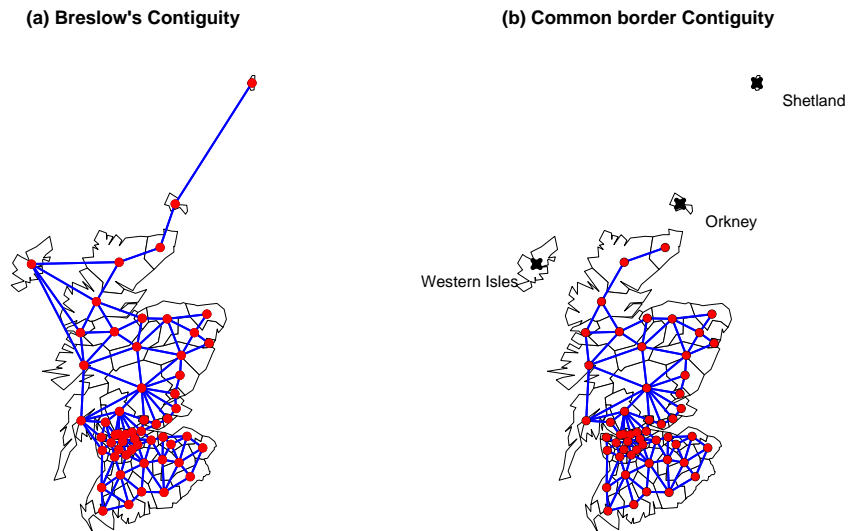


Fig. 5.2. Neighbouring structure for Scottish data: (a) Contiguity defined in Breslow and Clayton (1993), (b) contiguity based on sharing a common border.

applying these CAR models to irregular lattices, the imposed neighborhood structure and the spatial correlation could be misleading and strongly dependent to the number of neighbours. Furthermore the neighborhood criteria must be sometimes carefully examined. For instance, in the case of very irregular regions with different sizes and shapes or in the presence of not contiguous regions like islands.

Figure 5.2 shows two different adjacency matrix for the Scottish data set. Considering the common border criteria, the isles of Shetland, Orkney and Western Isles have no neighbours and the total number of neighbours is 234. We used the adjacency matrix defined by Breslow and Clayton (1993) in order to fit the CAR models with a number of 264 neighbours. We show the results obtained with this last matrix since it is the one commonly used in the literature. However, it is worth mentioning that results were different depending on the neighborhood criteria used.

For the CAR models, the parameter  $\phi$  presents high values ( $\approx 1$ ) for both models (4.3) and (4.4), which denotes that all the variation is explained by the spatial autocorrelation. For the intrinsic Besag's CAR model, similar interpre-

tation can be obtained for the estimated parameters, the spatial correlation structure absorbs the overdispersion without variability in each region, and the three alternative CAR formulations present similar results on model parameters.

Table 5.1 shows the better performance of Smooth-CAR models in terms of AIC and BIC criteria. As we mentioned above, it is important to notice that, although the AIC and BIC values of the Negative Binomial are smaller than the ones obtained in other models, they cannot be compared, since the distribution assumed for the data is different. In Figure 5.3 we can see both, large geographical trend and local spatial variation. If we compare Figure 5.3 with Figure 5.1, we can see that in the Smooth-CAR model the large-scale trend is smoother than in the PRIDE model. This could be expected since in the PRIDE model all the spatial variation is fitted by the  $P$ -spline. The partition of the spatial variation seems more realistic in Figure 5.3. However, more research is still needed to check to what extent it is possible to separate both spatial effects, or whether we could really only look at the overall fit.

A residual analysis was also performed for both PRIDE and Smooth-CAR models. The spatial deviance residuals exhibit a small-scale spatial dependence in the data not captured by the models. This dependency appears located around the high populated urban areas like Glasgow, Dundee or Edinburgh.

## 6 Discussion

Through the paper we have presented different approaches to the analysis of spatial count data, based on the combination of two-dimensional  $P$ -splines and individual random effects.

The mixed model representation of multidimensional  $P$ -splines for spatial coordinates, using the properties of the “Box-product” yields a fully parametric



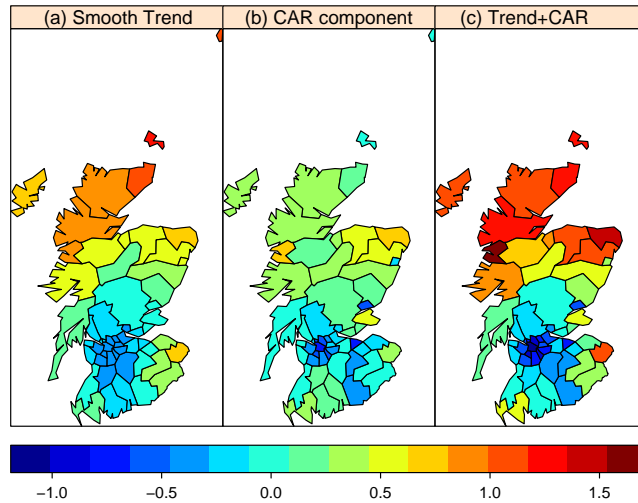


Fig. 5.3. Smooth-CAR model: (a) Smooth Trend ( $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ ), (b) CAR structured random effects (**b**) and (c) the sum of trend and CAR component.

model which can be easily implemented in standard statistical software. In this sense, the smoothing and spatial correlation given by the spatial random effects can be estimated simultaneously in REML and PQL context. The usual spatial models consider the spatial correlation between two locations  $s_i$  and  $s_j$  as  $\text{Corr}[\mathbf{s}_i, \mathbf{s}_j] = \rho_{ij}$ , where  $\rho$  is a function of the distance between centroids of areas  $i$  and  $j$ . This type of model assumes the same correlation in all spatial directions. When considering environmental problems where climate factors are important, the assumption of isotropy is questionable. We give an alternative by using anisotropic smoothers. The smoothing parameters  $\boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_2$ , permit different amount of smoothing in each spatial coordinate, through the penalty matrix  $\mathbf{F}$  in (2.13). This suppose an advantage on the interpretation of the spatial correlation structure, which with other models is not so straightforward to understand.

The Penalized Random Individual Dispersion Effects (PRIDE) model proposed by Perperoglou and Eilers (2007) was applied to spatial counts in the

mixed model context allowing for individual regional effects. Alternatively, Negative Binomial distribution was also considered.

Finally, we propose the combination of the  $P$ -spline smooth approach with conditionally autoregressive models. Spatial smoothing is used to model the large-scale variability as an overall trend, and the small-scale or local neighbour-based heterogeneity is accounted for by a CAR structure. The analysis of Scottish lip cancer data showed the attractiveness of this new approach.

Further simulation work needs to be done in order to test the identifiability of both smooth and CAR effects, since there might be situations when the Smooth-CAR model is not identifiable. Another important issue to be addressed is dependence of the fitted model on the arbitrariness of the neighbourhood criteria (e.g. when the areas are of different shapes or sizes or when there are discontinuous regions as isles) and the assignment of the centroids of the spatial locations as a summary of the whole region considered. A further extension of these models will consider the incorporation of relevant covariates or more complex structures.

## References

- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B*, 36:192–236.
- Besag, J. (1984). A method for the analysis of field experiments based on first differences. In *Spatial Methods in Field Experiments. Proceedings of the Biometric Society Workshop*, pages 9–11, University of Durham.
- Besag, J. and Green, P. J. (1993). Spatial statistics and bayesian computation. *J. R. Statist. Soc. B*, 55(1):25–37.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.

- Besag, J., York, J. C., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43:1–59.
- Breslow, N. E. and Clayton, D. G. (1993). Aproximated inference in generalised linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.
- Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Assoc.*, 93:961–994.
- Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Number 30 in Econometric Society Monograph. Cambridge University Press.
- Clayton, D. G. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–682.
- Congdon, P. (2006). A model for non-parametric spatially varying regression effects. *Computational Statistics and Data Analysis*, 50:422–445.
- Congdon, P. (2007). Mixtures of spatial and unstructured effects for spatially discontinuous health outcomes. *Computational Statistics and Data Analysis*, 51:3197–3212.
- Cressie, N. (1993). *Statistics for Spatial Data (Revised Edition)*. John Wiley and Sons, Inc., New York.
- Cressie, N. and Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406):393–401.
- Currie, I. D. and Durbán, M. (2002). Flexible smoothing with  $P$ -splines: A unified approach. *Statistical Modelling*, 4:333–349.
- Currie, I. D., Durbán, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.
- Currie, I. D., Durbán, M., and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc.*

- B*, 68:1–22.
- Dean, C., Ugarte, M. D., and Militino, A. F. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57:197–202.
- Dean, C. B. and Lawless, J. F. (1989). Tests for detecting overdispersion in poisson regression models. *Journal of the American Statistical Association*, 84(406):467–472.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics*, 47:299–350.
- 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(1):61–76.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with *B*-splines and penalties. *Stat. Sci.*, 11:89–121.
- Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: A bayesian perspective. *Statistica Sinica*, 14:715–745.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability. Chapman and Hall, London.
- Hinde, J. and Demetrio, C. (1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27:151–170.
- Hurvich, C. and Simonoff, J. (1998). Smoothing parameter selection in non-parametric regression using an improved akaike information criterion. *J. R. Statist. Soc. B*, 60:271–293.
- Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society, C - Applied Statistics*, 52:1–18.
- Kelsall, J. and Wakefield, J. (2002). Modelling spatial variation in disease risk: A geostatistical approach. *Journal of the American Statistical Association*, 97(459):692–701.

- Kemp, I., Boyle, P., Smans, M., and Muir, C. (1985). *Atlas of Cancer in Scotland, 1975-1980: Incidence and Epidemiologic Perspective*. IARC Scientific Publication 72, Lyon, France, International Agency for Research on Cancer.
- Kneib, T. and Fahrmeir, L. (2006). Structured additive regression for categorical space-time data: A mixed model approach. *Biometrics*, 62:109–118.
- Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.
- Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics*, 15:209–225.
- Leroux, B. G., Lei, X., and Breslow, N. (1999). *Estimation of disease rates in small areas: a new mixed model for spatial dependence*. Statistical models in Epidemiology, The Environment, and Clinical Trials. Springer, New York.
- Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *J. Roy. Stat. Soc., B*, 61:381–400.
- Liu, S. (1999). Matrix results on the khatri-rao and tracy-singh products. *Linear Algebra and its Applications*, 289:267–277.
- Marx, B. D. and Eilers, P. H. C. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis*, 28:193–209.
- Militino, A. F., Ugarte, M. D., and Dean, C. B. (2001). The use of mixture models for identifying high risks in disease mapping. *Statistics in Medicine*, 20(13):2035–2049.
- Patterson, H. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.
- Perperoglou, A. and Eilers, P. H. C. (2007). Overdispersion modelling with individual random effects and penalized likelihood (submitted).
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, UK. ISBN: 0521785162.

- Speed, T. (1991). Comment on “BLUP is a good thing: The estimation of random effects”, by Robinson, G. K. *Stat. Sci.*, 6:15–51.
- Thurston, S., Wand, M., and Wiencke, J. (2000). Negative binomial additive models. *Biometrics*, 56:139–144.
- Verbyla, A., Cullis, B., Kenward, M., and Welham, S. (1999). The analysis of designed experiments and longitudinal data using smoothing splines. *J. Roy. Stat. Soc. C*, 48:269–312.
- Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–183.
- Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, 18:223–249.
- Welham, S. J., Cullis, B. R., Kenward, M. G., and Thompson, R. (2007). A comparison of mixed model splines for curve fitting. *Aust. N. Z. J. Stat.*, 49(1):1–23.
- Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.
- Yasui, Y. and Lele, S. (1997). A regression method for spatial disease rates: An estimating function approach. *Journal of the American Statistical Association*, 92(437):21–32.