Ayma, D., Durbán, M., Lee, D. J. & Eilers, P. H. (2016). Penalized composite link models for aggregated spatial count data: A mixed model approach. *Spatial Statistics*, vol. 17, pp. 179–198.

DOI: 10.1016/j.spasta.2016.06.007

Penalized composite link models for aggregated spatial count data: a mixed model approach

Diego Ayma[a], María Durbán[a], Dae-Jin Lee[b,*], Paul H. C. Eilers[c]

[a] *Department of Statistics, Universidad Carlos III de Madrid, 28911 Leganés, Spain.*
[b] *Basque Center for Applied Mathematics, 48009 Bilbao, Spain.*
[c] *Department of Biostatistics, Erasmus University Medical Center, 3015GE Rotterdam, The Netherlands.*

Abstract

Mortality data provide valuable information for the study of the spatial distribution of mortality risk, in disciplines such as spatial epidemiology and public health. However, they are frequently available in an aggregated form over irregular geographical units, hindering the visualization of the underlying mortality risk. Also, it can be of interest to obtain mortality risk estimates on a finer spatial resolution, such that they can be linked to potential risk factors that are usually measured in a different spatial resolution. In this paper, we propose the use of the penalized composite link model and its mixed model representation. This model considers the nature of mortality rates by incorporating the population size at the finest resolution, and allows the creation of mortality maps at a finer scale, thus reducing the visual bias resulting from the spatial aggregation within original units. We also extend the model by considering individual random effects at the aggregated scale, in order to take into account the overdispersion. We illustrate our novel proposal using two datasets: male lip cancer incidence in Scotland counties, and female deaths by lung cancer in Indiana, USA. We also compare the performance of our proposal with the area-to-point Poisson kriging approach.

Keywords: Penalized composite link models, Mixed models, Mortality rates, Disease mapping

1. Introduction

Disease mapping studies has to deal with public health data that are frequently only available in an aggregated form over irregular geographical units,

*Corresponding author
Email addresses: dayma@est-econ.uc3m.es (Diego Ayma), mdurban@est-econ.uc3m.es (María Durbán), dlee@bcamath.org (Dae-Jin Lee), p.eilers@erasmusmc.nl (Paul H. C. Eilers)

like counties, districts, and municipalities. Epidemiologists, health care practitioners, and other researchers use these data to study the spatial distribution of mortality risk (caused by a certain disease), and thus identify areas of excess and their potential risk factors. In general, rates are used as measures of the risk, since they incorporate information about the population of each unit. Choropleth maps are then used to display such rates, but they must be interpreted with caution: rates calculated from small or sparsely populated units are likely to be elevated artificially (Waller and Gotway, 2004). This effect, known as the "small number problem", may hinder the detection of meaningful patterns in the study area. Another problem that can arise is the spatial misalignment between potential risk factors and health data: in general, the former are available on a finer spatial resolution than the latter. For example, most deprivation indices are built on the smallest possible geographical units of a certain region (see Rey et al., 2009; Salmond and Crampton, 2012) or even on a fine grid (Caudeville et al., 2012). Environmental agents (such as air pollution) constitute examples of risk factors that vary continuously in space. Consequently, this issue precludes their direct use in a correlation analysis, which is a critical step for disease control intervention. Therefore, it is relevant to develop spatial methodologies that filter the noise caused by the small number problem and allow the creation of mortality maps, from aggregated data, at a finer spatial resolution.

Different approaches have been used to reduce the noise in aggregated mortality rates (see Besag et al., 1991; MacNab and Dean, 2002; Fahrmeir et al., 2004; Goovaerts, 2005; Lee and Durbán, 2009; among others). However, they give smoothed mortality estimates that are assumed constant over each unit, yielding a coarse spatial trend. To obtain a more detailed impression of mortality through units, several methodologies have been proposed in the literature. In a geostatistical framework, Kelsall and Wakefield (2002) obtained pointwise posterior medians of the underlying continuous risk surface, for colorectal cancer mortality in the UK district of Birmingham, via a Gaussian random field (GRF) model. Goovaerts (2006) generalized the Poisson kriging algorithm given by Monestiez et al. (2005, 2006), which incorporates the size and shape of the units, as well as the population density, into the filtering of noisy mortality rates. This generalization allows the mapping of the corresponding mortality risk at a fine resolution. The performance of his approach, called area-to-point Poisson kriging, was compared with two geostatistical methods. The first one corresponds to the simple interpolation of raw rates to the nodes of a fine grid using ordinary kriging. The second one corresponds to the approach proposed by Berke (2004), in which the raw rates are replaced by their global empirical Bayes estimates before the interpolation process. Local Bayes estimates were also considered in the analysis. Lately, and from a Bayesian inferential viewpoint, Diggle et al. (2013) used the class of log-Gaussian Cox processes (as models for spatial point process data) to construct a continuous map of lung cancer mortality risks in the Castile-La Mancha region of Spain, from spatially discrete data. The previous works are related to the "change of support problem" (see, e.g., Gotway and Young, 2002), since they seek to obtain mortality

risk estimates at a fine resolution from data available at coarse geographical units. There has been substantial work on this problem, especially within the hierarchical Bayesian modelling approach (see Mugglin and Carlin, 1998; Zhu et al., 2000; Mugglin et al., 2000; Gelfand et al., 2001; Banerjee et al., 2015, Ch. 7; among others).

In this paper, we propose the use of the Penalized Composite Link Model (PCLM, Eilers, 2007) for the case of spatial aggregation, together with its mixed model representation. This model, which we call throughout the paper Composite Link Mixed Model, allows us to create mortality risk maps from aggregated data at a fine spatial resolution, and to incorporate finer scale information into the filtering of noisy mortality rates. We assume here the underlying mortality risk at the fine resolution is smooth. The flexibility of the model is provided by the use of B-splines, together with a penalty on the regression coefficients, following the P-spline methodology (Eilers and Marx, 1996). The mixed model representation makes it possible to include specific random effects or further correlation structure if necessary, and to estimate the parameters of the PCLM under the framework of mixed model theory.

We illustrate the case when we seek to estimate the spatial mortality trend at a fine grid, using health data available at coarse geographical units, i.e., the area-to-point (ATP) case. We obtain a continuous surface without spatial boundaries on the study area (that is, an isopleth map), reducing the visual bias associated with the interpretation of choropleth maps (Cressie, 1993) that is caused by the variation in shape and size of the units.

The rest of this paper is organized as follows. In Section 2, we present our methodological approach: the Composite Link Mixed Model (or more briefly, CLMM) for spatially aggregated count data, where we indicate how the ATP case is accommodated by our proposal. Also, in this section we provide a parameter estimation procedure for the CLMM, and we extend the model to deal with the problem of overdispersion in count data, by incorporating individual random effects at the aggregated scale. In Section 3, we apply our methodology using two datasets related with female deaths by lung cancer in the state of Indiana (USA), recorded over the period 1970-1994, and with male deaths by lip cancer in Scotland, recorded over the period 1975-1980. We specifically use the Scottish lip cancer dataset to illustrate how our model accommodates the presence of overdispersion in data. In Section 4, we use the lung cancer dataset to compare the performance of our proposal with the area-to-point Poisson kriging of Goovaerts (2006) (an additional performance comparison, where the geographical units vary considerably in shape and size, is included in Appendix A). Finally, we provide a short discussion in Section 5.

2. The composite link mixed model

2.1. The PCLM approach

In the one-dimensional case, suppose that a vector of aggregated counts $\boldsymbol{y}$ follows a Poisson distribution with mean vector $\boldsymbol{\mu}$. These counts can be seen as

indirect observations of a latent process that we want to estimate. The penalized composite link model approach of Eilers (2007) (which is based on the work by Thompson and Baker, 1981) offers an elegant way to do this, by considering $\boldsymbol{\mu}$ as composed of latent expectations. The Poisson PCLM is given by:

$$\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma} = \mathbf{C}\exp(\mathbf{B}\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\gamma}$ represents the mean vector of the latent process at a desirable fine resolution, $\mathbf{C}$ is the composition matrix that describes how these latent expectations are combined to yield $\boldsymbol{\mu}$, $\mathbf{B} = \mathbf{B}(\boldsymbol{x})$ is a B-spline basis constructed from a covariate, $\boldsymbol{x}$, at fine resolution, and $\boldsymbol{\theta}$ is the associated vector of regression coefficients. Smoothness is imposed over adjacent regression coefficients, by subtracting the roughness penalty $\frac{1}{2}\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$ from the log-likelihood of $\boldsymbol{y}$, where $\mathbf{P} = \lambda\mathbf{D}'\mathbf{D}$ is based on a difference matrix $\mathbf{D}$ of order q and a smoothing parameter λ that controls the amount of smoothness. Parameter estimation of the model given in Eq. (1), subject to the penalization, is carried out by a modified version of the iteratively reweighted least squares (IRWLS) algorithm, where an information criterion (such as AIC or BIC) is used to choose an optimal value for λ. Several applications of the PCLM can be found in Eilers (2012).

For illustration purposes, let us consider the number of deaths from respiratory disease of American population in January 1959, from ages 1 to 120 (see Currie et al., 2006, for more details about these data). Figure 1 shows the counts per age-at-death (vertical lines) and the smooth trend that follow these counts ($g = 1$). If we artificially aggregate them into two, five, ten, and twenty age classes, and we apply the PCLM approach to these aggregated counts, we obtain the smooth colored curves of Figure 1 ($g = 2, 5, 10, 20$). The smooth curves for the cases $g = 2$, $g = 5$ and $g = 10$ are close to the smooth trend at the disaggregated scale, whereas the blue smooth curve for the case $g = 20$ departs from it (especially between 60 and 90 years old). This is because we have less precision when the aggregation level increases.

In a Bayesian framework, the PCLM approach was developed by Lambert and Eilers (2009) for the estimation of smooth densities from grouped data. Furthermore, Lambert (2011) extended the previous work for the estimation of a bivariate density from a histogram with wide bins, using semi- and non-parametric strategies. In that case, Kronecker versions of the composition and B-splines matrices are considered. However, its extension for the spatial setting has not been explored.

In the following subsection, we extend the PCLM in Eq. (1) to the spatial case. The development of this extension allows to analyse the distribution of mortality rates (disease incidence, fertility or others vital rates) in a finer spatial resolution than the original, under the modest assumption of smoothness. This implies an improvement over previous related works (Lee and Durbán, 2009; Perperoglou and Eilers, 2010) in terms of the visualization of the underlying mortality risk (the previous cited works only provide mortality risk estimates for each unit while our approach enables to create a mortality risk surface across coarse units) and the incorporation of fine-scale information in the mortality

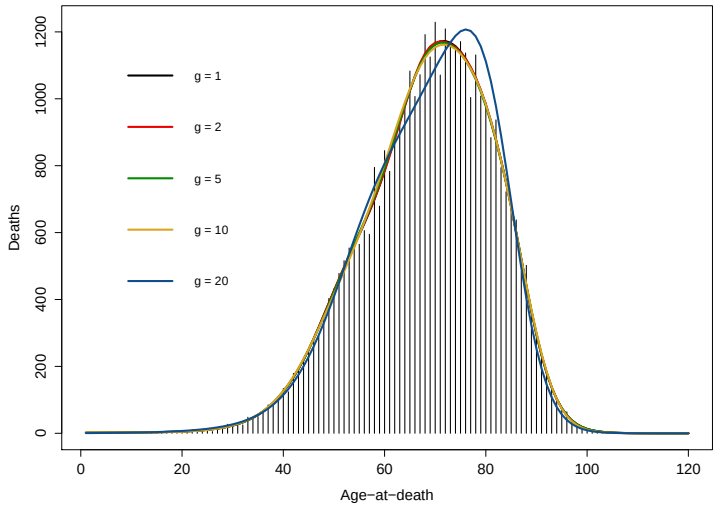

Figure 1: Death counts from respiratory disease of American population in January 1959, from ages 1 to 120 (vertical lines). The black curve represents the estimated trend based on the ungrouped data. The colored curves represent the estimated distributions using the PCLM approach, from different aggregations per g age classes, where g denotes the width of the groups.

risk estimation. Also we choose to represent the PCLM as a mixed model. This representation is useful because it allows us to include specific random effects or further correlation structure, and offers another alternative for the parameter estimation of the PCLM – avoiding the use of information criteria for the smoothing parameter selection.

2.2. The spatial CLMM

Suppose the vector of aggregated counts $\boldsymbol{y}$ are now available over n non-overlapping geographical units $\boldsymbol{v}_i$, $i = 1, ..., n$. Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be the geographical coordinates (longitude and latitude, respectively) of length m that define the desirable fine spatial resolution. Then, in this new context, the full regression basis $\mathbf{B}$ is defined as the "row-wise" Kronecker product (denoted by $\square$, Eilers et al., 2006) of the marginal B-spline bases $\mathbf{B}_1 = \mathbf{B}(\boldsymbol{x}_1)$ and $\mathbf{B}_2 = \mathbf{B}(\boldsymbol{x}_2)$ of dimension $m \times c_1$ and $m \times c_2$, respectively:

$$\mathbf{B} = \mathbf{B}_2 \square \mathbf{B}_1 = (\mathbf{B}_2 \otimes \mathbf{1}'_{c_1}) \odot (\mathbf{1}'_{c_2} \otimes \mathbf{B}_1), \tag{2}$$

where $\mathbf{1}_k$ denotes a vector of ones of length k, and the matrix operators $\otimes$ and $\odot$ represent the Kronecker and the Hadamard (or "element-wise") products, respectively. The construction of $\mathbf{B}_1$ and $\mathbf{B}_2$ depends on the number of selected (equally-spaced) knots for each coordinate, ndx_1 and ndx_2, and the degree of

the B-splines used, $bdeg_1$ and $bdeg_2$. The two-dimensional penalty matrix is given by:

$$\mathbf{P} = \lambda_1 \mathbf{I}_{c_2} \otimes \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 \otimes \mathbf{I}_{c_1}, \tag{3}$$

where $\mathbf{I}_k$ is an identity matrix of dimension $k \times k$, λ_d is the smoothing parameter that controls the amount of smoothness along the covariate $\boldsymbol{x}_d$, and $\mathbf{P}_d = \mathbf{D}_d' \mathbf{D}_d$ is the marginal penalty matrix based on the difference matrix $\mathbf{D}_d$ of order q_d ($d = 1, 2$). The penalty matrix in Eq. (3) allows for anisotropy, i.e., we can have a different amount of smoothing for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Here we have to make choices about ndx_d, $bdeg_d$, and q_d, for $d = 1, 2$. Usually, ndx_1 and ndx_2 are chosen large (one knot for every four different covariate values is a reasonable choice) to preserve enough flexibility. For the other quantities, it is sufficient to use cubic B-splines (that is, $bdeg_1 = bdeg_2 = 3$) and quadratic penalties ($q_1 = q_2 = 2$). For a further discussion, see Eilers and Marx (1996), Currie and Durbán (2002) and Eilers et al. (2015).

Considering the regression basis in Eq. (2) and its associated regression coefficients $\boldsymbol{\theta}$, it was shown in Lee and Durbán (2009) that expression $\mathbf{B}\boldsymbol{\theta}$ can be reformulated as $\mathbf{B}\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ (using a suitable orthogonal transformation matrix $\mathbf{T}$ such that $\mathbf{BT} = [\mathbf{X} : \mathbf{Z}]$ and $\mathbf{T}'\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})'$) where $\mathbf{X}$ and $\mathbf{Z}$ are the fixed and random effects matrices, and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are their associated coefficients, respectively. The construction of the mixed model matrices $\mathbf{X}$ and $\mathbf{Z}$ is briefly described below (for more details, see Lee and Durbán, 2009 and Lee, 2010, pp. 63-65).

Consider the singular value decomposition (SVD) of the marginal penalty matrix $\mathbf{P}_d$ in Eq. (3):

$$\mathbf{P}_d = \mathbf{U}_d \boldsymbol{\Sigma}_d \mathbf{U}_d',$$

where $\mathbf{U}_d$ is the matrix of singular vectors, and $\boldsymbol{\Sigma}_d$ is a diagonal matrix that contains the singular values of $\mathbf{P}_d$, for $d = 1, 2$. Each matrix $\mathbf{U}_d$ can be split in two parts:

$$\mathbf{U}_d = [\mathbf{U}_{dn} : \mathbf{U}_{ds}],$$

where $\mathbf{U}_{ds}$ is a matrix of dimension $c_d \times (c_d - q_d)$ that contains the non-null part of the decomposition. With this partition, we can decompose each marginal penalty matrix as follows:

$$\mathbf{P}_d = [\mathbf{U}_{dn} : \mathbf{U}_{ds}] \begin{bmatrix} \mathbf{O}_{q_d} & \\ & \tilde{\boldsymbol{\Sigma}}_d \end{bmatrix} [\mathbf{U}_{dn} : \mathbf{U}_{ds}]',$$

where $\mathbf{O}_{q_d}$ denotes a square matrix of zeroes of dimension $q_d \times q_d$, and $\tilde{\boldsymbol{\Sigma}}_d$ is a diagonal matrix that contains the $(c_d - q_d)$ positive singular values of $\mathbf{P}_d$, for $d = 1, 2$. Then, defining the matrices $\mathbf{X}_d = \mathbf{B}_d \mathbf{U}_{dn}$ and $\mathbf{Z}_d = \mathbf{B}_d \mathbf{U}_{ds}$ ($d = 1, 2$), the mixed model matrices are obtained as:

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_2 \square \mathbf{X}_1, \\ \mathbf{Z} &= [\mathbf{Z}_2 \square \mathbf{X}_1 : \mathbf{X}_2 \square \mathbf{Z}_1 : \mathbf{Z}_2 \square \mathbf{Z}_1]. \end{aligned} \tag{4}$$

Moreover, due to the transformation matrix $\mathbf{T}$ and the penalty matrix given in Eq. (3), it can be shown the mixed model penalty corresponds to the block-diagonal matrix:

$$\mathbf{F} = \begin{bmatrix} \lambda_2\tilde{\mathbf{\Sigma}}_2 \otimes \mathbf{I}_{q_1} & & \\ & \lambda_1\mathbf{I}_{q_2} \otimes \tilde{\mathbf{\Sigma}}_1 & \\ & & \lambda_2\tilde{\mathbf{\Sigma}}_2 \otimes \mathbf{I}_{c_1-q_1} + \lambda_1\mathbf{I}_{c_2-q_2} \otimes \tilde{\mathbf{\Sigma}}_1 \end{bmatrix}, \quad (5)$$

where the matrices $\tilde{\mathbf{\Sigma}}_d$ $(d = 1, 2)$ were defined above.

Using the previous mixed model representation, we can extend the model given in Eq. (1) to the spatial case by modifying $\boldsymbol{\gamma}$ as follows:

$$\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma} = \mathbf{C}\exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \log(\boldsymbol{e}_{\mathrm{f}})), \text{ with } \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \quad (6)$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the mixed model matrices defined in Eq. (4), and $\boldsymbol{e}_{\mathrm{f}}$ is a vector of exposures at the fine resolution. The random effects have covariance matrix $\mathbf{G}$, which is obtained as $\mathbf{G} = \sigma^2_\epsilon\mathbf{F}^{-1}$, where $\sigma^2_\epsilon = 1$ (in the Poisson case) and $\mathbf{F}$ is the penalty matrix defined in Eq. (5). We refer to Eq. (6) as the (Poisson) composite link mixed model or, more briefly, as CLMM.

Since our goal is to analyse rates, the vector $\boldsymbol{e}_{\mathrm{f}}$ in Eq. (6) has to be known in advance; otherwise, it has to be estimated. If the vector of exposures is only available at the aggregated level, a naive approach to estimate $\boldsymbol{e}_{\mathrm{f}}$ is to assume that these aggregated exposures are evenly distributed throughout the fine resolution. Another possibility is to apply the CLMM approach to the aggregated vector of exposures to obtain estimates for $\boldsymbol{e}_{\mathrm{f}}$.

The composition matrix $\mathbf{C}$ in Eq. (6) is fixed and its structure depends on the process that generates the aggregated data. If we take $\mathbf{C}$ as the identity matrix, then we have that $\boldsymbol{\mu} = \boldsymbol{\gamma}$ in Eq. (6). In such case, the CLMM approach is reduced to the Penalized Generalized Linear Mixed Model (PGLMM) approach of Lee and Durbán (2009) for Poisson data, where the spatial covariates correspond to the geographical centroids of the units.

In Section 1, we pointed out the CLMM can handle the ATP case. For that, we consider $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as the coordinates of the cell centroids of a fine grid, which fall inside of the geographical units. Thus, the elements of the associated (spatial) composition matrix $\mathbf{C}$ become:

$$c_{ij} = \begin{cases} 1 & \text{if } (x_{1j}, x_{2j}) \text{ belongs to unit } \boldsymbol{v}_i \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

for $i = 1, ..., n$, and $j = 1, ..., m$.

2.3. Parameter estimation

Since the covariance matrix $\mathbf{G}$ in Eq. (6) is determined from the block-diagonal matrix $\mathbf{F}$ in Eq. (5), it depends on two smoothing parameters, λ_1 and λ_2, that have to be estimated. In consequence, the parameter estimation of the CLMM involves two interrelated stages: (a) the estimation of the fixed and random coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of the vector of latent expectations $\boldsymbol{\gamma}$; and (b) the

estimation of the smoothing parameters λ_1 and λ_2. For stage (a), we use the penalized quasi-likelihood (PQL) approach (Breslow and Clayton, 1993), which is commonly used for the parameter estimation of GLMMs; and for stage (b), we use the residual (or restricted) maximum log-likelihood (REML, Patterson and Thompson, 1971) as a numerical optimization criterion for the smoothing parameter selection. Technical details of these stages are derived below.

Consider the joint density function of $\boldsymbol{y}$ in the CLMM context:

$$f(\boldsymbol{y}|\boldsymbol{\alpha}) = \exp\left(\boldsymbol{y}'\log(\boldsymbol{\mu}) - \mathbf{1}_n'\boldsymbol{\mu} - \mathbf{1}_n'\log(\Gamma\left(\boldsymbol{y}+\mathbf{1}_n\right))\right), \tag{8}$$

where $\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma}$, $\boldsymbol{\gamma} = \exp(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\alpha}+\log(\boldsymbol{e}_{\mathrm{f}}))$, and $\boldsymbol{\alpha} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{G}(\lambda_1, \lambda_2)\right)$. Taking into account Eq. (8) and for given values of λ_1 and λ_2, we obtain estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by maximizing the penalized log-likelihood:

$$\ell_{\mathrm{p}} = \log\left(f(\boldsymbol{y}|\boldsymbol{\alpha})\right) - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{G}^{-1}\boldsymbol{\alpha}. \tag{9}$$

Differentiating Eq. (9) with respect to β_k and α_l, we obtain:

$$\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_k} = \sum_{i=1}^{n}\left((y_i-\mu_i)\frac{1}{\mu_i}\sum_{j=1}^{m} c_{ij}\gamma_j x_{jk}\right), \text{ for } k = 1, ..., p; \tag{10}$$

$$\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_l} = \sum_{i=1}^{n}\left((y_i-\mu_i)\frac{1}{\mu_i}\sum_{j=1}^{m} c_{ij}\gamma_j z_{jl}\right) - \mathbf{G}_l^{-1}\boldsymbol{\alpha}, \text{ for } l = 1, ..., r, \tag{11}$$

where $\mathbf{G}_l^{-1}$ denotes the l-th row of the matrix $\mathbf{G}^{-1}$. Writing $\frac{1}{\mu_i}\sum_{j=1}^{m} c_{ij}\gamma_j x_{jk}$ in Eq. (10) and $\frac{1}{\mu_i}\sum_{j=1}^{m} c_{ij}\gamma_j z_{jl}$ in Eq. (11) as $\breve{x}_{ik}$ and $\breve{z}_{il}$, respectively, and equating the expressions above to zero, we obtain:

$$\sum_{i=1}^{n}(y_i-\mu_i)\breve{x}_{ik} = 0, \text{ for } k = 1, ..., p; \tag{12}$$

$$\sum_{i=1}^{n}(y_i-\mu_i)\breve{z}_{il} = \mathbf{G}_l^{-1}\boldsymbol{\alpha}, \text{ for } l = 1, ..., r. \tag{13}$$

Moreover, Eq. (12) and Eq. (13) can be rewritten in matrix form as:

$$\breve{\mathbf{X}}'\left(\boldsymbol{y}-\boldsymbol{\mu}\right) = \mathbf{0}; \tag{14}$$

$$\breve{\mathbf{Z}}'\left(\boldsymbol{y}-\boldsymbol{\mu}\right) = \mathbf{G}^{-1}\boldsymbol{\alpha}, \tag{15}$$

where $\breve{\mathbf{X}} = \mathbf{W}^{-1}\mathbf{C}\boldsymbol{\Gamma}\mathbf{X}$ and $\breve{\mathbf{Z}} = \mathbf{W}^{-1}\mathbf{C}\boldsymbol{\Gamma}\mathbf{Z}$, with $\mathbf{W} = \mathrm{diag}(\boldsymbol{\mu})$ and $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. Defining the working vector:

$$\boldsymbol{z} = \breve{\mathbf{X}}\boldsymbol{\beta} + \breve{\mathbf{Z}}\boldsymbol{\alpha} + \mathbf{W}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}\right),$$

the solution of Eq. (14) and Eq.(15) via Fisher scoring algorithm (see Green, 1987) can be expressed as the iterative solution of the system:

$$\begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{X}} & \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{Z}} \\ \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{X}} & \mathbf{G}^{-1}+\breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{Z}} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\boldsymbol{z} \\ \breve{\mathbf{Z}}'\mathbf{W}\boldsymbol{z} \end{bmatrix}. \tag{16}$$

Notice that the linear system given in Eq. (16) has exactly the same structure as that for a PGLMM (Lee, 2010). The difference is that in a PGLMM we would have $\mathbf{X}$ and $\mathbf{Z}$ while here $\breve{\mathbf{X}}$ and $\breve{\mathbf{Z}}$ appear. Thus $\breve{\mathbf{X}}$ and $\breve{\mathbf{Z}}$ are "working" $\mathbf{X}$ and $\mathbf{Z}$ matrices, respectively. From Eq. (16) we obtain a modified version of the standard mixed model estimators:

$$\widehat{\boldsymbol{\beta}} = (\breve{\mathbf{X}}'\mathbf{V}^{-1}\breve{\mathbf{X}})^{-1}\breve{\mathbf{X}}'\mathbf{V}^{-1}\boldsymbol{z}, \tag{17}$$

$$\widehat{\boldsymbol{\alpha}} = \mathbf{G}\breve{\mathbf{Z}}'\mathbf{V}^{-1}(\boldsymbol{z} - \breve{\mathbf{X}}\widehat{\boldsymbol{\beta}}), \tag{18}$$

where:

$$\mathbf{V} = \mathbf{W}^{-1} + \breve{\mathbf{Z}}\mathbf{G}\breve{\mathbf{Z}}'. \tag{19}$$

Conditioning on the estimates given in Eq. (17) and Eq. (18), the smoothing parameters λ_1 and λ_2 can be estimated numerically by maximizing the residual maximum log-likelihood (REML):

$$-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\breve{\mathbf{X}}'\mathbf{V}^{-1}\breve{\mathbf{X}}| - \frac{1}{2}\boldsymbol{z}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\breve{\mathbf{X}}(\breve{\mathbf{X}}'\mathbf{V}^{-1}\breve{\mathbf{X}})^{-1}\breve{\mathbf{X}}'\mathbf{V}^{-1})\boldsymbol{z}. \tag{20}$$

Therefore, the PQL solution is achieved by iteration between Eq. (17), Eq. (18), and Eq. (20), until convergence.
Once the parameter values at convergence are obtained, we can derive standard errors for the mixed model estimators as shown in Lin and Zhang (1999), i.e. by approximating the covariance matrix of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})'$ by its Bayesian counterpart. This approximated covariance matrix is given by:

$$\mathbf{M} = \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{X}} & \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{Z}} \\ \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{X}} & \mathbf{G}^{-1} + \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{Z}} \end{bmatrix}^{-1}, \tag{21}$$

which corresponds to the inverse of the matrix on the left-hand side of Eq. (16). Thus we can obtain standard errors for $\widehat{\boldsymbol{\eta}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{\alpha}}$ by taking the square root of the diagonal elements of $\mathbb{V}ar(\widehat{\boldsymbol{\eta}})$, which are obtained as:

$$\mathbb{V}\mathrm{ar}(\widehat{\eta}_j) = \mathrm{diag}([\mathbf{X} : \mathbf{Z}]\mathbf{M}[\mathbf{X} : \mathbf{Z}]')_{jj},$$

where $\mathbf{M}$ is defined in Eq. (21). Approximate standard errors for $\exp(\widehat{\boldsymbol{\eta}})$ can be derived by using the Delta method (see, e.g., Ver Hoef, 2012; Agresti, 2015):

$$\mathbb{V}\mathrm{ar}(\exp(\widehat{\eta}_j)) = \mathbb{V}\mathrm{ar}(\widehat{\eta}_j) \times (\exp(\widehat{\eta}_j))^2.$$

The effective dimension (ED) of the CLMM (on the aggregated scale) is the trace of the so-called "hat matrix", $\mathrm{tr}(\mathbf{H})$, as shown in Hastie and Tibshirani (1990), which is given by:

$$\mathbf{H} = [\breve{\mathbf{X}} : \breve{\mathbf{Z}}]\mathbf{M} \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W} \\ \breve{\mathbf{Z}}'\mathbf{W} \end{bmatrix},$$

with $\mathbf{M}$ defined in Eq. (21). Then, we can calculate the Akaike Information Criterion (AIC) as:

$$\text{AIC} = \text{Dev}(\boldsymbol{y}|\hat{\boldsymbol{\mu}}) + 2 \times \text{ED} \ ,$$

where $\text{Dev}(\boldsymbol{y}|\hat{\boldsymbol{\mu}})$ is the deviance for the Poisson case given by:

$$\text{Dev}(\boldsymbol{y}|\hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left(y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right)\right).$$

2.4. Overdispersion

The PCLM approach is a useful tool for modelling aggregated or grouped counts. However, it is assumed the counts for the groups follow Poisson distributions. When this is not the case, because of overdispersion (i.e., the presence of extra Poisson variation due to an unobserved heterogeneity), underestimation of the variability of estimates may occur. As a solution for the overdispersion problem, we propose to introduce individual random effects for the logarithms of the expected values, one for each group count. This can be viewed as an adaptation of the PRIDE ("Penalized Regression with Individual Deviance Effects") approach given by Perperoglou and Eilers (2010) and Lee and Durbán (2009). Here, we develop this idea under the CLMM framework; thus we will refer to this approach throughout the paper as CLMM-P.

Consider $\boldsymbol{\phi} = \mathbf{C}\boldsymbol{\gamma}$, where $\mathbf{C}$ is the composition matrix and $\boldsymbol{\gamma}$ is the vector of latent expectations at the fine resolution, with $\boldsymbol{\gamma} = \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \log(\boldsymbol{e}_{\text{f}}))$. We can generalize the CLMM formulation by assuming that the aggregated counts are now Poisson distributed with mean vector:

$$\boldsymbol{\mu} = \exp(\log(\boldsymbol{\phi}) + \boldsymbol{\delta}),\ \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})\ ,\ \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \kappa^{-1}\mathbf{I}_n), \tag{22}$$

where κ is the dispersion parameter associated with the individual random effects $\boldsymbol{\delta}$. These random effects (defined at the aggregated scale) provides a device to absorb the overdispersion that causes the extra-variability. Thus, in the model given by Eq. (22), we are simultaneously dealing with parameters at aggregated and at a finer scale.

Considering the penalized log-likelihood:

$$\ell^*_{\text{p}} = \log\left(f(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\delta})\right) - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{G}^{-1}\boldsymbol{\alpha} - \frac{1}{2}\kappa\boldsymbol{\delta}'\boldsymbol{\delta},$$

where $f(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\delta})$ denotes the joint density distribution of $\boldsymbol{y}$ in the CLMM-P context, and using the PQL approach for the estimation of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta}$ in Eq. (22), we obtain the following system of equations:

$$\begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{X}} & \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{Z}} & \breve{\mathbf{X}}'\mathbf{W} \\ \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{X}} & \mathbf{G}^{-1} + \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{Z}} & \breve{\mathbf{Z}}'\mathbf{W} \\ \mathbf{W}\breve{\mathbf{X}} & \mathbf{W}\breve{\mathbf{Z}} & \kappa\mathbf{I}_n + \mathbf{W} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix} = \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\boldsymbol{z} \\ \breve{\mathbf{Z}}'\mathbf{W}\boldsymbol{z} \\ \mathbf{W}\boldsymbol{z} \end{bmatrix}, \tag{23}$$

where now the "working" matrices are defined as $\breve{\mathbf{X}} = \mathbf{\Phi}^{-1}\mathbf{C}\mathbf{\Gamma}\mathbf{X}$ and $\breve{\mathbf{Z}} = \mathbf{\Phi}^{-1}\mathbf{C}\mathbf{\Gamma}\mathbf{Z}$, with $\mathbf{\Phi} = \text{diag}(\boldsymbol{\phi})$ and $\mathbf{\Gamma} = \text{diag}(\boldsymbol{\gamma})$. In this case, the matrix of weights and the working vector are $\mathbf{W} = \text{diag}(\boldsymbol{\mu})$, with $\boldsymbol{\mu}$ defined as in Eq. (22), and $\boldsymbol{z} = \breve{\mathbf{X}}\boldsymbol{\beta} + \breve{\mathbf{Z}}\boldsymbol{\alpha} + \boldsymbol{\delta} + \mathbf{W}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$, respectively.

It is possible to reduce the large system of equations given in Eq. (23) by defining $\boldsymbol{\delta}$ as:

$$\boldsymbol{\delta} = (\mathbf{W} + \kappa\mathbf{I}_n)^{-1}\mathbf{W}(\boldsymbol{z} - \breve{\mathbf{X}}\boldsymbol{\beta} - \breve{\mathbf{Z}}\boldsymbol{\alpha}). \tag{24}$$

Thus, if we define:

$$\mathbf{W}^* = \kappa(\mathbf{W} + \kappa\mathbf{I}_n)^{-1}\mathbf{W},$$

we have that $\kappa\boldsymbol{\delta} = \mathbf{W}^*(\boldsymbol{z} - \breve{\mathbf{X}}\boldsymbol{\beta} - \breve{\mathbf{Z}}\boldsymbol{\alpha})$. Using this result in Eq. (23), we obtain:

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

This leads to the same system of equations of the Poisson CLMM without overdispersion (see Eq. (16)), but changing the matrix of weights to $\mathbf{W}^*$ and the addition of the vector $\boldsymbol{\delta}$ to the working vector. Therefore, the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are estimated as in Eq. (17) and Eq. (18), with $\mathbf{V} = \mathbf{W}^{*-1} + \breve{\mathbf{Z}}\mathbf{G}\breve{\mathbf{Z}}'$, and $\boldsymbol{\delta}$ is estimated using Eq. (24). Then, conditioning on these estimates, the smoothing parameters λ_1 and λ_2 and the dispersion parameter κ are estimated by Eq. (20).

Finally, to compute AIC, the "hat-matrix" in this case is given by:

$$\mathbf{H}^* = [\breve{\mathbf{X}} : \breve{\mathbf{Z}} : \mathbf{I}_n] \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{X}} & \breve{\mathbf{X}}'\mathbf{W}\breve{\mathbf{Z}} & \breve{\mathbf{X}}'\mathbf{W} \\ \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{X}} & \mathbf{G}^{-1} + \breve{\mathbf{Z}}'\mathbf{W}\breve{\mathbf{Z}} & \breve{\mathbf{Z}}'\mathbf{W} \\ \mathbf{W}\breve{\mathbf{X}} & \mathbf{W}\breve{\mathbf{Z}} & \kappa\mathbf{I}_n + \mathbf{W} \end{bmatrix}^{-1} \begin{bmatrix} \breve{\mathbf{X}}'\mathbf{W} \\ \breve{\mathbf{Z}}'\mathbf{W} \\ \mathbf{W} \end{bmatrix}.$$

For an efficient calculation of the trace of the hat matrix, see Perperoglou and Eilers (2010).

3. Applications

In this section, we apply our methodology using two real mortality datasets. We use the first dataset to illustrate the CLMM approach for the ATP case. With the second dataset, we illustrate how the CLMM-P approach can handle the problem of overdispersion, often present in count data. For parameter estimation, we follow the methodology described in Section 2.3, where we use the L-BFGS-B method of Byrd et al. (1995) to optimize the REML log-likelihood given in Eq. (20). For both datasets, we also compare our methodology with the ATP Poisson kriging of Goovaerts (2006). Hereafter we refer to this approach as PK.

We implemented the CLMM and CLMM-P approaches in the statistical software `R` version 3.1.0 (64-bit), and a 3.40 GHz Intel® Core™ i7 processor computer with 4 GB of RAM and Windows® 7 operating system. The PK approach is implemented in the geostatistical software SpaceStat 4.0 (`http://www.biomedware.com/`).

3.1. Lung cancer dataset

The lung cancer dataset comes from the Atlas of Cancer Mortality in the United States (Pickle et al., 1999) and was downloaded from `http://ratecalc.cancer.gov`. This dataset has been previously analysed by Goovaerts (2006) and it contains the number of white female deaths by lung cancer and the corresponding age-adjusted mortality rates (per 100000 person-years), recorded over the period 1970-1994 in the state of Indiana at county level (92 counties in total). The population-at-risk in each county can be estimated with the formula:

$$\frac{\text{Total number of deaths (1970-1994)}}{\text{Age-adjusted mortality rate (1970-1994)}} \times 100000 \ .$$

Goovaerts (2006) imposed a 55×94 grid (with grid cells of 25 km^2) over the map of Indiana, leading to 3751 grid points that fall inside the map (see Figure 2a). Next, he allocated the previous county-level population estimates to this fine grid, according to the 2000 census block level data. Figure 2b shows the spatial distribution of the population-at-risk on the fine grid, which reflects the heterogeneous repartition of population in Indiana. These high-resolution population estimates were kindly provided by Dr. Pierre Goovaerts (BioMedware Inc., MI, USA) and we will use them in subsequent analysis.

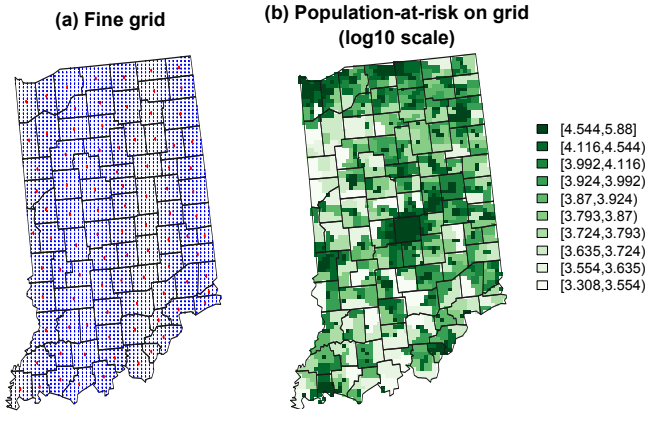

Figure 2: The left map shows the fine grid obtained by imposing a 55×94 grid over the map of Indiana, leading to 3751 grid points (blue points). The right map shows the spatial distribution of the population-at-risk for this fine grid (using a log10 scale).

Figure 3a shows the spatial distribution of age-adjusted mortality rates (per 100000 person-years) for lung cancer in Indiana. We use a yellow-red color scheme for data visualization, where the class boundaries correspond to the deciles of the original rates. Rates higher than the median tend to be more red as they depart from it, while lower rates tend to be more yellow. Since the sizes of the counties in Indiana are relatively similar, it is easy to identify areas of

excess in this region. The highest rates are reported for the counties of Clark (30.637), Johnson (30.726), and Marion (31.624), which is the most populated county in Indiana.

To reduce the noise present in lung cancer mortality rates, we first apply the PGLMM approach (Lee and Durbán, 2009) with the spatial coordinates of the county centroids as covariates, second order penalties, and 22 equally-spaced knots (applying around one knot for every four different centroid coordinates) for each marginal cubic B-spline basis. The resulting smoothed mortality risk is shown in Figure 3b, with range varying from 13.302 to 31.624. The maximum rates after smoothing are still located in counties with the highest lung cancer rates. This situation was pointed out also by Goovaerts (2006), when he analysed these data (at county level) with different kriging methods. For this dataset, if we increase the number of knots in the PGLMM approach, we will obtain a similar spatial mortality pattern to that shown in Figure 3b.

Now we apply the CLMM approach on this dataset to obtain a continuous mortality risk map. To do that, let us consider the number of white female deaths by lung cancer per county as the vector of aggregated counts ($\boldsymbol{y}$), and the population-at-risk on the fine grid of 25 km^2 cells (displayed in Figure 2b) as the vector of exposures at fine resolution ($\boldsymbol{e}_{\mathrm{f}}$). To set up the CLMM formulation, we use the spatial coordinates of the grid points (see Figure 2a) as covariates at fine resolution, second order penalties, and 22 equally-spaced knots for each marginal cubic B-spline basis. Then, we can construct the spatial composition matrix as is described in Eq. (7). Figure 3c shows the resulting CLMM mortality risk, which is calculated as $\widehat{r}_{\mathrm{CLMM}} = 100000 \times \exp(\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{\alpha}})$. Regarding the computing time, the estimation process took a little more than 1 minute. This isopleth map gives a more detailed impression of the mortality distribution, where areas of lower and higher mortality risks are clearly delineated on the map of Indiana. Higher risk estimates are still observed in the counties of Clark, Johnson, and Marion, while lower risk estimates are more concentrated in some south-western and north-eastern counties of Indiana.

To compare our proposal with other existing methods, we apply the PK approach of Goovaerts (2006) on this dataset. Given the fine grid point $\boldsymbol{u}_j = (x_{1j}, x_{2j})$, $j = 1, ..., 3751$, within a geographical unit $\boldsymbol{v}_\delta$, the PK estimator is obtained as a linear combination of the kernel rate $r(\boldsymbol{v}_\delta)$ and the rates observed in $(K-1)$ neighboring units:

$$\widehat{r}_{\mathrm{PK}}(\boldsymbol{u}_j) = \sum_{i=1}^{K} \lambda_i(\boldsymbol{u}_j) r(\boldsymbol{v}_i),$$

where $\lambda_i(\boldsymbol{u}_j)$ is the weight assigned to the rate $r(\boldsymbol{v}_i)$ when estimating the risk at $\boldsymbol{u}_j$. The K kriging weights are computed by solving a system of linear equations, in which the weights $\lambda_i(\boldsymbol{u}_j)$ are constrained to sum up 1, and a point-support covariance of the risk, or equivalently a point-support semivariogram $\gamma_{\mathrm{R}}(\boldsymbol{h})$ is required to solve it. Since only aggregated data are available, this function cannot be estimated directly from the observed rates. Goovaerts (2008) developed a procedure to conduct the derivation of $\gamma_{\mathrm{R}}(\boldsymbol{h})$ from the "regularized" experi-

mental semivariogram computed from areal data (i.e., "deconvolution" process), in presence of irregular geographical units and heterogeneous population distribution.

Figure 3d shows the resulting PK mortality risk, using the software indicated above, together with the indications given in Goovaerts (2006) for the estimation of this continuous surface. The PK approach provides a similar spatial pattern to the CLMM approach, with some discrepancies in the north and south-east of the central counties. We should note that the application of both approaches on this dataset produces some risk estimates at fine scale that exceed the maximum raw lung mortality rate (31.795). For example, the maximum risk estimates for the CLMM and PK are 34.067 and 33.896, respectively.

Figure 4 shows the standard error maps associated with the mortality risk estimates given at the bottom of Figure 3. The PK standard errors are calculated as the square root of the PK variances (see Goovaerts, 2006, Eq. 12). Most of the CLMM standard errors are lower than those obtained with the PK, through Indiana counties, showing that CLMM reduces the uncertainty.

To compare the aggregations resulting from the CLMM and PK approaches, we can compute the corresponding AIC using the estimated means $\hat{\boldsymbol{\mu}}_{\text{CLMM}}$ and $\hat{\boldsymbol{\mu}}_{\text{PK}}$, respectively. The first one is calculated as in Eq. (6) while the elements of the second are obtained from Goovaerts (2006, Eq. 15) as:

$$\hat{\mu}_{\text{PK}}(\boldsymbol{v}_i) = 10^{-5} \times e(\boldsymbol{v}_i)\widehat{r}_{\text{PK}}(\boldsymbol{v}_i) = 10^{-5} \times \sum_{j=1}^{P_i} e(\boldsymbol{u}_j) r(\boldsymbol{u}_j), \tag{25}$$

where P_i denotes the number of grid points $\boldsymbol{u}_j$ used to discretize the county $\boldsymbol{v}_i$, and $e(\boldsymbol{v}_i) = \sum_{j=1}^{P_i} e(\boldsymbol{u}_j)$, for $i = 1, ..., n$. The resulting AIC for the CLMM and PK (at county level) are 163.565 and 237.394, respectively.

In order to assess the prediction performance among the mentioned approaches, we have carried out a simulation study in Section 4.

3.2. Scottish lip cancer dataset

The Scottish lip cancer dataset (Clayton and Kaldor, 1987) has been widely analysed in the literature. In this section, we apply the CLMM-P approach (developed in Section 2.4) on this dataset, to obtain a continuous surface that take into account the overdispersion present in count data.

This dataset consists of the observed ($\boldsymbol{y}$) and expected ($\boldsymbol{e}$) number of male cases of lip cancer, recorded in 56 counties in Scotland over the period 1975-1980. Figure 5a shows the spatial distribution of the Standardized Mortality Rates (SMR) on a logarithmic scale for lip cancer mortality, which is obtained as:

$$\log(\text{SMR}) = \log\left(\frac{y_i}{e_i}\right), \text{ for } i = 1, ..., 56.$$

We see that most of the higher raw log(SMR) are located in the north of Scotland; specifically in the counties of Caithness, Ross and Cromarty, Skye and Lochalsh, and Banff and Buchan.

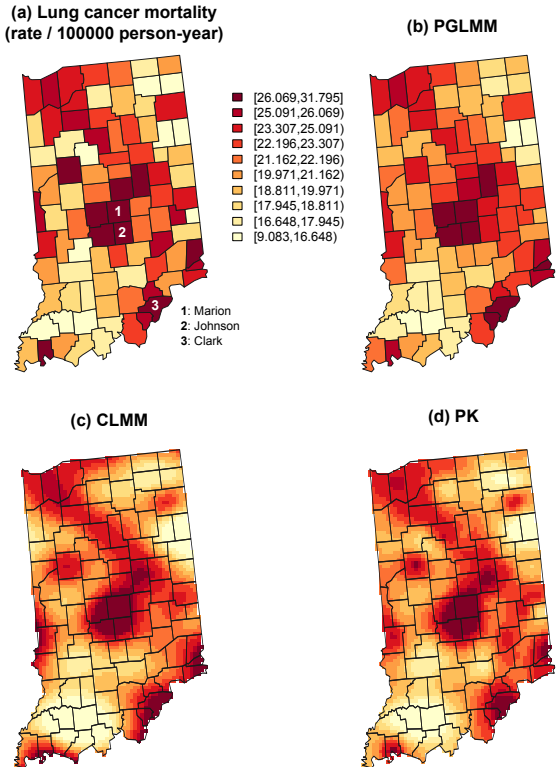

Figure 3: Map of lung cancer mortality rates in Indiana, and the risk estimated by different approaches. The top-left map displays the age-adjusted mortality rates per 100000 person-years recorded over the period 1970-1994, and the top-right map shows the smoothed mortality risks resulting from the PGLMM approach. The bottom maps show the smoothed mortality risks estimated using the CLMM (bottom-left) and PK (bottom-right) approaches. The color legend applies to all maps; the class boundaries correspond to the deciles of the original rates.

In order to apply the CLMM approach, we impose a 120×120 fine grid over the map of Scotland, leading to 3855 grid points that fall inside the map. Since we lack of the vector of exposures at this fine scale, we estimate them using the naive approach described in Section 2.2. We denote this vector as $\hat{\boldsymbol{e}}_{\text{naive}}$. To set up the CLMM formulation, we use 25 equally-spaced knots for each marginal cubic B-spline basis and second order penalties. Then, the corresponding spatial composition matrix is constructed as is described in Eq. (7). Figure 5b shows the resulting CLMM estimates for the log(SMR) at the selected fine grid (that is, $\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\alpha}}$). Regarding the computing time, the estimation process took a little less than 2 minutes. From Figure 5b, we observe there exist an increasing trend from the more central counties to the ones of the coast, and also from south to

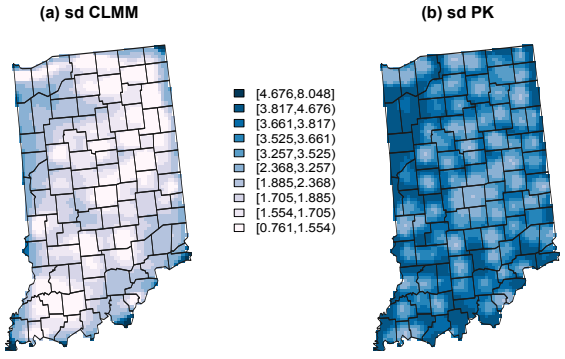

Figure 4: Standard error maps for lung cancer mortality risks in Indiana, estimated by (a) CLMM and (b) PK approaches.

north. Moreover, using the previous point estimates, we can obtain a smooth trend for the log(SMR) at county level by aggregate them in the following way:

$$\log(\text{SMR})_{\text{CLMM}} = \log\left(\frac{\hat{\boldsymbol{\mu}}_{\text{CLMM}}}{\boldsymbol{e}}\right), \tag{26}$$

where $\hat{\boldsymbol{\mu}}_{\text{CLMM}}$ is obtained as in Eq. (6), with $\boldsymbol{e}_{\text{f}} = \hat{\boldsymbol{e}}_{\text{naive}}$. Figure 5c shows these coarse estimates for the log(SMR) at county level.

Now we apply the CLMM-P approach on this dataset. For that we use the same settings as in the CLMM approach. Figure 5d shows the resulting CLMM-P estimates for the log(SMR) at the selected fine grid, where we have included the estimated individual random effects, $\hat{\boldsymbol{\delta}}$, at the fine scale to take into account the overdispersion. This is done by adding the term $\mathbf{C}^{-}\hat{\boldsymbol{\delta}}$ to the estimated spatial trend (that is, $\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\alpha}} + \mathbf{C}^{-}\hat{\boldsymbol{\delta}}$), where $\mathbf{C}^{-}$ denotes the Moore-Penrose inverse of $\mathbf{C}$. This matrix can be easily computed as $\mathbf{C}^{-} = \left(\mathbf{C}\mathbf{R}^{-1}\right)'$, where $\mathbf{R}$ is a diagonal matrix whose elements are the sums of the rows of $\mathbf{C}$. Regarding the computing time, the estimation process took a little more than 2 minutes. We observe that this map presents some differences with respect to the map obtained by CLMM approach, especially in the north of Scotland. Similarly to what we did before, we can obtain a smooth trend for the log(SMR) at county level, from the CLMM-P estimates, as:

$$\log(\text{SMR})_{\text{CLMM-P}} = \log\left(\frac{\hat{\boldsymbol{\mu}}_{\text{CLMM-P}}}{\boldsymbol{e}}\right), \tag{27}$$

where $\hat{\boldsymbol{\mu}}_{\text{CLMM-P}}$ is obtained as in Eq. (22), with $\boldsymbol{e}_{\text{f}} = \hat{\boldsymbol{e}}_{\text{naive}}$. These coarse estimates for log(SMR) at county level are displayed in Figure 5e.

To compare our proposal with other existing methods, we apply the PK approach on this dataset. Figure 5f shows the resulting estimates for log(SMR)

at the selected fine grid using this approach, which is similar to that produced by the CLMM approach. Also, we can obtain a smooth trend for the log(SMR) at county level, from the PK estimates, as:

$$\log(\text{SMR})_{\text{PK}} = \log\left(\frac{\hat{\boldsymbol{\mu}}_{\text{PK}}}{\boldsymbol{e}}\right), \tag{28}$$

where $\hat{\boldsymbol{\mu}}_{\text{PK}}$ is obtained as in (25). These coarse estimates for log(SMR) at county level are displayed in Figure 5g.

Figure 6 shows the standard error maps associated with the mortality risk estimates given (in log scale) at the middle of Figure 5. In this case, we observe that higher errors are located in the islands of the north and north-west of Scotland. In these parts, the errors associated to the CLMM and CLMM-P approaches are greater than those associated with the PK approach. The higher standard errors in CLMM and CLMM-P approaches might be due to the presence of the islands where there is a discontinuity in the boundaries (the tensor product smooth tends to interpolate the sea where no data are available leading to larger standard errors), while PK model implemented in Spacestat 4.0 uses an areal deconvolution process with the definition of a spatial weight matrix with a minimum distance to ensure that all units will be connected with at least one other unit (Jacquez et al., 2014). Some advances in spline smoothing can be studied to include special penalties to account for smoothing in complex and irregular domains (see Ramsay, T., 2002; Wood et al., 2008).

In order to compare the aggregations resulting from the CLMM, CLMM-P and PK approaches, we can compute the AIC using the estimated means $\hat{\boldsymbol{\mu}}_{\text{CLMM}}$, $\hat{\boldsymbol{\mu}}_{\text{CLMM-P}}$ and $\hat{\boldsymbol{\mu}}_{\text{PK}}$ already calculated in Eq. (26)-(28), respectively. The resulting AIC for the CLMM, CLMM-P and PK (at county level) are 110.8, 89.8, and 186.7, respectively, showing that the CLMM-P is more appropriate in presence of overdispersion.

4. Simulation study

In this section we perform a simulation study to compare the prediction performance of CLMM approach with the ATP Poisson kriging (PK) of Goovaerts (2006). For that, we use the lung cancer dataset described in Section 3.1.

The simulation study was conducted in the following way:

1. The continuous mortality surface obtained with the PK approach was considered here as the true underlying mortality trend over the fine grid of 25 km^2 cells in Indiana. We denoted these mortality rates as $r(\boldsymbol{u}_j)$, $j = 1, ..., 3751$, where $\boldsymbol{u}_j$ represent the coordinates of the fine grid points.
2. These quantities and the population-at-risk over each 25 km^2 cell of the fine grid (denoted as $e(\boldsymbol{u}_j)$) were used to calculate the mortality rate for each county $\boldsymbol{v}_i$, $i = 1, ..., 92$:

$$r(\boldsymbol{v}_i) = \frac{1}{e(\boldsymbol{v}_i)} \sum_{j=1}^{P_i} e(\boldsymbol{u}_j) r(\boldsymbol{u}_j),$$

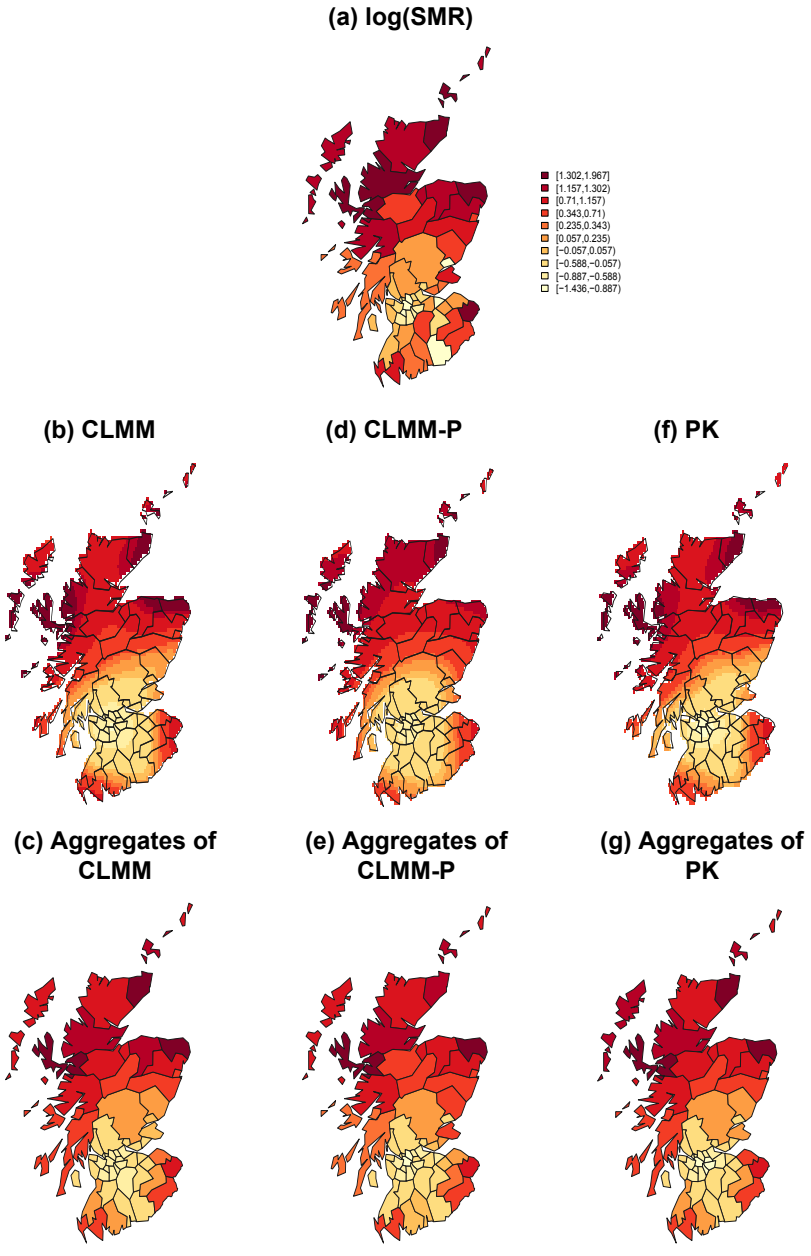

Figure 5: Map of (log) standardized mortality rates in Scotland, and the (log) mortality risks estimated by different approaches. The top map shows the log(SMR) recorded over the period 1975-1980 for 56 counties. The middle maps show the smoothed (log) mortality risks at a selected fine grid, using CLMM, CLMM-P, and PK. The bottom maps show the resulting aggregation of these point estimates. The color legend applies to all maps; the class boundaries correspond to the deciles of the log(SMR).

where P_i denotes the number of points $\boldsymbol{u}_j$ used to discretize the county $\boldsymbol{v}_i$, and $e(\boldsymbol{v}_i) = \sum_{j=1}^{P_i} e(\boldsymbol{u}_j)$.

3. 100 realizations of the number of deaths recorded over each county were generated by random drawing of a Poisson distribution whose mean parameter is $r(\boldsymbol{v}_i) \times e(\boldsymbol{v}_i)$.
4. For each realization, we apply the CLMM and PK approaches, using the population-at-risk over the fine grid of 25 km^2 cells as the vector $\boldsymbol{e}_{\mathrm{f}}$ of exposures at the fine resolution.

For all $l = 1, ..., 100$ realizations, the predicted risks $r_{\mathrm{P}}^{(l)}(\boldsymbol{u}_j)$ obtained from both approaches were compared to the underlying risk $r(\boldsymbol{u}_j)$, $j = 1, ..., 3751$, using the following criteria:

- Mean error (ME):

$$\mathrm{ME}^{(l)} = \frac{1}{W} \sum_{j=1}^{3751} e(\boldsymbol{u}_j) \left(r_{\mathrm{P}}^{(l)}(\boldsymbol{u}_j) - r(\boldsymbol{u}_j) \right) \text{ with } W = \sum_{j=1}^{3751} e(\boldsymbol{u}_j)$$

- Mean absolute error (MAE):

$$\mathrm{MAE}^{(l)} = \frac{1}{W} \sum_{j=1}^{3751} e(\boldsymbol{u}_j) \left| r_{\mathrm{P}}^{(l)}(\boldsymbol{u}_j) - r(\boldsymbol{u}_j) \right| \text{ with } W = \sum_{j=1}^{3751} e(\boldsymbol{u}_j)$$

- Root mean squared error (RMSE):

$$\mathrm{RMSE}^{(l)} = \sqrt{\frac{1}{W} \sum_{j=1}^{3751} e(\boldsymbol{u}_j) \left(r_{\mathrm{P}}^{(l)}(\boldsymbol{u}_j) - r(\boldsymbol{u}_j) \right)^2} \text{ with } W = \sum_{j=1}^{3751} e(\boldsymbol{u}_j)$$

In all these criteria, the prediction error at each grid point $\boldsymbol{u}_j$ is weighted according to the population size at that location. This was done to penalize more the errors that affect a larger population (Goovaerts, 2006). Notice that for the ME criterion, it could happen that positive and negative errors are canceled out so that the true error is underestimated. We have included ME criteria in order to follow the same comparisons as in Goovaerts (2005).

Figure 7 shows these resulting errors via box-plots, in which we observe that our approach gives better prediction accuracy than the PK approach, for each criterion. Table 1 gives the averages and the standard deviations of the resulting errors (for each criterion) derived from the simulation study. Notice that these results are obtained from a region where the geographical units (the counties) are similar in shape and size. We have conducted an additional simulation study, in which the units vary greatly in shape and size (see Appendix A). For that case, we have considered the Scottish lip cancer dataset. This new simulation study shows how the performance of the composite link mixed model is also satisfactory for irregular geographical units.

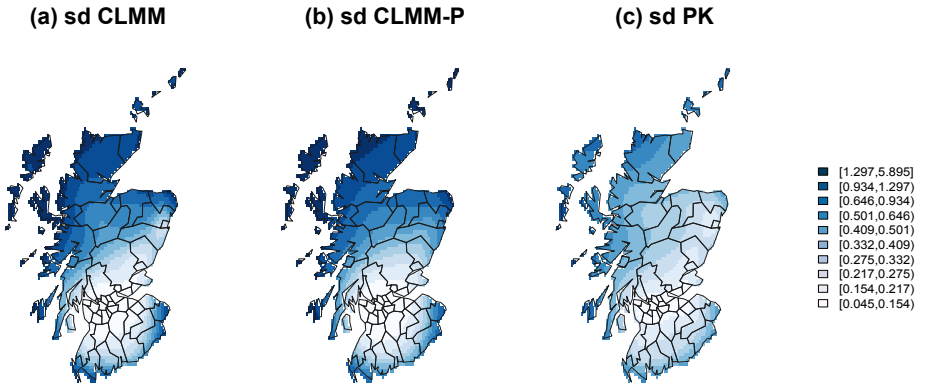

Figure 6: Standard error maps for lip cancer mortality risks in Scotland, estimated by (a) CLMM, (b) CLMM-P and (c) PK approaches.

Approach	ME		MAE		RMSE	
	avg	std	avg	std	avg	std
CLMM	0.0000	0.0006	0.9687	0.0005	1.2553	0.0006
PK	0.0062	0.0005	1.0197	0.0011	1.3514	0.0013

Table 1: Performance comparison of CLMM and PK approaches, using different criteria: mean errors (ME), mean absolute errors (MAE), and root mean squared errors (RMSE). These errors are summarized in terms of the average (avg) and standard deviation (std).

5. Discussion

We presented and applied the composite link mixed model for spatially aggregated data to the disaggregation of mortality rates. It provides a flexible descriptive tool for epidemiological studies, when the aim is to visualize the spatial distribution of certain rates at a desirable spatial resolution. The CLMM approach filters the existing noise in raw rates, which is caused by the small number problem, and allows the creation of more refined mortality maps by in-cluding the distribution of the exposure variable at fine resolution. The resulting CLMM estimates may be linked with potential risks factors that are available over the fine resolution, allowing a posterior correlation analysis between them. Under this framework, we included individual random effects at the aggregated scale to take into account the overdispersion problem, commonly occurring in count data. These individual random effects can be easily included at the fine scale (for graphical representation) by means of the Moore-Penrose inverse of the composition matrix. Since the CLMM is flexible, no assumptions about the covariance structure of the spatial process should be made (in contrast to kriging methods). The penalty on the coefficients accounts for estimating the spatial trend and the amount of smoothing on each longitude and latitude dimensions. For irregular domains (such as it was the case of the northern counties and the

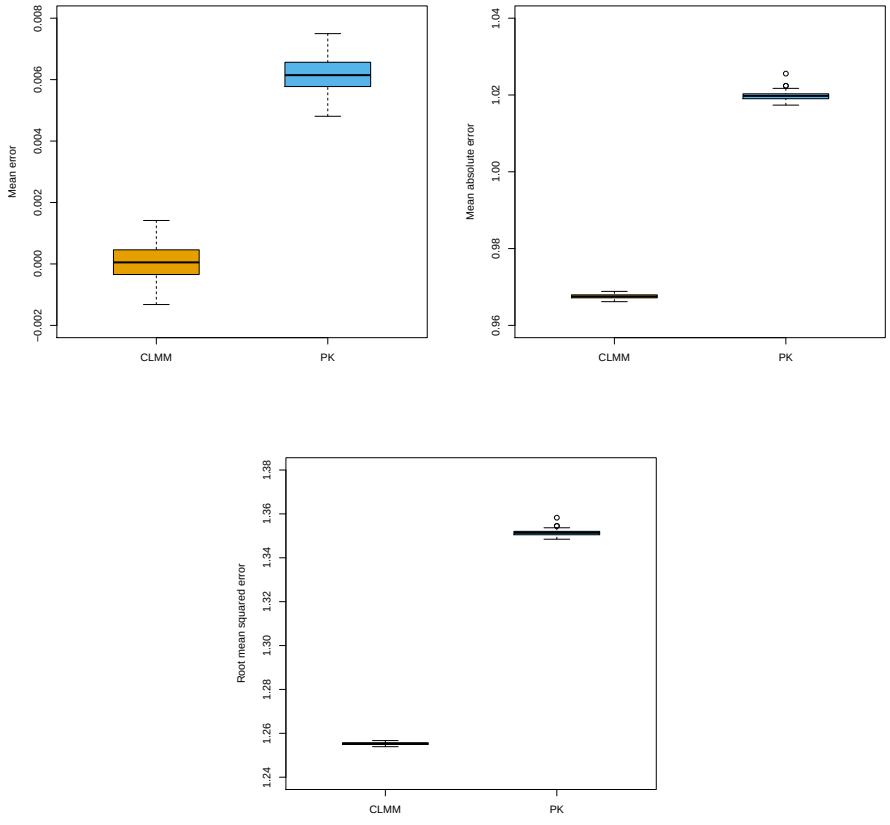

Figure 7: Performance comparison between CLMM and PK approaches using different criteria: mean errors (top-left), mean absolute errors (top-right), and root mean squared errors (bottom).

presence of discontinuities or islands) a possible solution in the CLMM approach is the use of special penalties over complex domains as in Wood et al. (2008) where smoothers are designed to not smooth across boundary features.

We used the statistical software `R` (R Core Team, 2015) for data analysis with the CLMM approach. Our plan is to implement the presented methodology in a future `R` package, in such a way that it can be accessible by any user. We provided some indications of computing time for the estimation of mortality trends at fine resolution in Section 3. Indeed, computation times will be reduced if we disaggregate at a lower spatial resolution (i.e. a less fine grid). A possibility to improve the computational speed is to accommodate the generalization of the Schall algorithm (Schall, 1991) presented by Rodríguez-Álvarez et al. (2015),

into a CLMM context.

The selected number of equally-spaced knots (for coordinates $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in Section 2.2) are chosen large, such that cover the study area. In the presented applications, if we increase the number of knots, we obtain similar continuous surfaces to those shown in this paper. On the contrary, if we choose very few knots, our approach will not be able to capture properly the underlying spatial trend behind aggregated data. For further details about the selection of knots in P-splines see Eilers and Marx, 1996, and Eilers et al., 2015.

We performed a simulation study to compare the area-to-point Poisson kriging of Goovaerts (2006) with our proposal, using aggregated data measured over the 92 counties of Indiana and the high-resolution population estimates over a fine grid. The simulation results showed that our proposal is competitive with respect to this geostatistical technique. An additional simulation study using the Scottish lip cancer dataset, where the counties greatly vary in shape and size, is detailed in Appendix A. Here while the accuracy of the CLMM model is better than PK, further research can be done to improve the smoothing in irregular domains.

Finally, and as future work, the proposed methodology can be generalized to the spatio-temporal setting, in which counts are also aggregated in time (i.e. by years or months). In this context, the implementation of efficient and fast algorithms for the estimation procedure of CLMMs will be critical. The resulting estimates will be displayed as dynamic maps, and will allow the comparison of mortality in the finest spatio-temporal resolution.

Acknowledgments

We would like to thank two reviewers and an associate editor for their constructive comments and suggestions on the original manuscript. We also thank Dr. Pierre Goovaerts, who provided the high resolution population estimates given in Section 3. This research was supported by the Spanish Ministry of Economy and Competitiveness grants MTM2011-28285-C02-02 and MTM2014-52184-P. The research of Dae-Jin Lee was also supported by the Basque Government through the BERC 2014-2017 and ELKARTEK programs and by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323. The research of Paul H. C. Eilers was also supported by the Universidad Carlos III de Madrid-Banco Santander Chair of Excellence program.

Appendix A. Simulation study of Scottish Lip Cancer dataset

In this appendix, we include an additional simulation study to compare the prediction performance among CLMM, CLMM-P and PK, when the geographical units vary considerably in shape and size. For that, we use the Scottish lip cancer dataset described in Section 3.2. Here we use the estimated vector of naive exposures as the true exposures at fine grid (that is, $\boldsymbol{e}_{\mathrm{f}} = \hat{\boldsymbol{e}}_{\mathrm{naive}}$).

The simulation study was conducted in a similar fashion as in Section 4, where the continuous mortality risk surface obtained with the PK approach was considered here as the true underlying mortality trend (see Figure 5f). Thus, for the resulting 100 realizations, the predicted risks $r_{\mathrm{P}}^{(l)}(\boldsymbol{u}_j)$ obtained from the three approaches were compared to the true underlying mortality risk, using the ME, MAE and RMSE criteria. Figure A.8 shows these resulting errors via boxplots, in which we observe the CLMM and CLMM-P approaches gives better prediction accuracy than PK, for each criterion. Note that, in this simulation setting, we did not include any overdispersion, and hence both CLMM and CLMM-P approaches are very similar. Table A.2 gives the averages and the standard deviations of the resulting errors (for each criterion) computed from this additional simulation study.

Approach	ME		MAE		RMSE	
	avg	std	avg	std	avg	std
CLMM-P	0.0040	0.0464	0.1523	0.0232	0.2748	0.0512
CLMM	0.0012	0.0463	0.1493	0.0216	0.2749	0.0505
PK	0.0552	0.0423	0.2041	0.0277	0.3191	0.0460

Table A.2: Performance comparison of CLMM-P, CLMM and PK approaches, using different criteria: mean errors (ME), mean absolute errors (MAE), and root mean squared errors (RMSE). These errors are summarized in terms of the average (avg) and standard deviation (std).

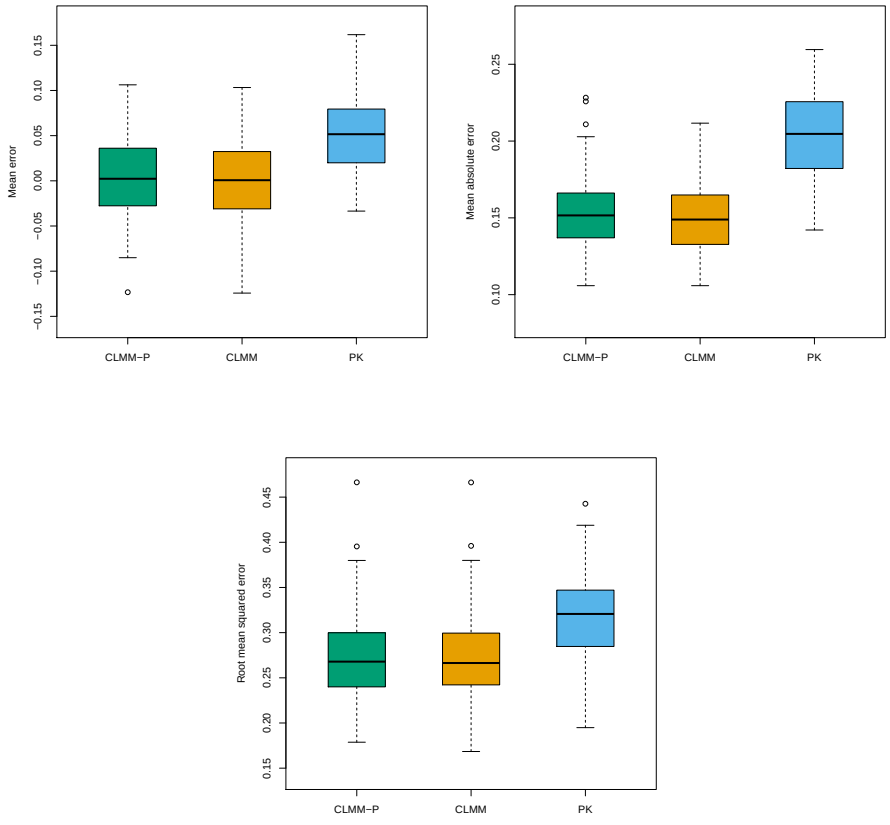

Figure A.8: Performance comparison between CLMM-P, CLMM and PK approaches using different criteria: mean errors (top-left), mean absolute errors (top-right), and root mean squared errors (bottom).

References

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models . John Wiley & Sons, New Jersey.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2015. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall, Boca Raton, Florida.

Berke, O., 2004. Exploratory disease mapping: kriging the spatial risk function from regional count data. International Journal of Health Geographics 3, 18. doi:10.1186/1476-072X-3-18.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two appli-

cations in spatial statistics. Annals of the Institute of Statistical Mathematics 43, 1–20. doi:10.1007/BF00116466.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9–25. doi:10.2307/2290687.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal of Scientific Computing 16, 1190–1208. doi:10.1137/0916069.

Caudeville, J., Bonnard, R., Boudet, C., Denys, S., Govaert, G., Cicolella, A., 2012. Development of a spatial stochastic multimedia exposure model to assess population exposure at regional scale. Science of The Total Environment 432, 297–308. doi:10.1016/j.scitotenv.2012.06.001.

Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in Disease Mapping. Biometrics 43, 671–681. URL: http://www.jstor.org/stable/2532003.

Cressie, N.A.C., 1993. Statistics for Spatial Data (revised edition). John Wiley & Sons, New York.

Currie, I.D., Durbán, M., 2002. Flexible smoothing with P-splines: a unified approach. Statistical Modelling 2, 333–349. doi:10.1191/1471082x02st039ob.

Currie, I.D., Durbán, M., Eilers, P.H.C., 2006. Generalized linear array models with applications to multidimensional smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68, 259–280. doi:10.1111/j.1467-9868.2006.00543.x.

Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M., 2013. Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. Statistical Science 28, 542–563. doi:10.1214/13-STS441.

Eilers, P.H.C., 2007. Ill-posed problems with counts, the composite link model and penalized likelihood. Statistical Modelling 7, 239–254. doi:10.1177/1471082X0700700302.

Eilers, P.H.C., 2012. Composite link, the neglected model, in: Komárek, A., Nagy, S. (Eds.), Proceedings of 27th International Workshop on Statistical Modelling, Prague, Czech Republic. pp. 11–23.

Eilers, P.H.C., Currie, I.D., Durbán, M., 2006. Fast and compact smoothing on large multidimensional grids. Computational Statistics & Data Analysis 50, 61–76. doi:10.1016/j.csda.2004.07.008.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11, 89–121.

Eilers, P.H.C., Marx, B.D., Durbán, M., 2015. Twenty years of P-splines. Statistics and Operations Research Transactions (SORT) 39, 149–186. URL: http://www.raco.cat/index.php/SORT/article/view/302258/391947.

Fahrmeir, L., Kneib, T., Lang, S., 2004. Penalized structured additive regression for space-time data: a Bayesian perspective. Statistica Sinica 14, 731–761.

Gelfand, A.E., Zhu, L., Carlin, B.P., 2001. On the change of support problem for spatio-temporal data. Biostatistics 2, 31–41. doi:10.1093/biostatistics/2.1.31.

Goovaerts, P., 2005. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. International Journal of Health Geographics 4, 31. doi:10.1186/1476-072X-4-31.

Goovaerts, P., 2006. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. International Journal of Health Geographics 5, 52. doi:10.1186/1476-072X-5-52.

Goovaerts, P., 2008. Kriging and semivariogram deconvolution in presence of irregular geographical units. Mathematical Geology 40, 101–128.

Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97, 632–648. doi:10.1198/016214502760047140.

Green, P.J., 1987. Penalized likelihood for general semi-parametric regression models. International Statistical Review 55, 245–259. doi:10.2307/1403404.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Adittive Models. Chapman and Hall, London.

Jacquez, G., Goovaerts, P., Kaufmann, A., Rommel, R., 2014. SpaceStat 4.0 User Manual: Software for the Space-Time Analysis of Dynamic Complex Systems. Fourth ed. Publisher: BioMedWare.

Kelsall, J., Wakefield, J., 2002. Modeling spatial variation in disease risk: a geostatistical approach. Journal of the American Statistical Association 97, 692–701. doi:10.1198/016214502388618438.

Lambert, P., 2011. Smooth semiparametric and nonparametric Bayesian estimation of bivariate densities from bivariate histogram data. Computational Statistics & Data Analysis 55, 429–445. doi:10.1016/j.csda.2010.05.011.

Lambert, P., Eilers, P.H.C., 2009. Bayesian density estimation from grouped continuous data. Computational Statistics & Data Analysis 53, 1388–1399. doi:10.1016/j.csda.2008.11.022.

Lee, D.-J., 2010. Smoothing mixed models for spatial and spatio-temporal data. Ph.D. thesis. Department of Statistics, Universidad Carlos III de Madrid, Spain.

Lee, D.-J., Durbán, M., 2009. Smooth-CAR mixed models for spatial count data. Computational Statistics & Data Analysis 53, 2968–2979. doi:10.1016/j.csda.2008.07.025.

Lin, X., Zhang, D., 1999. Inference in generalized additive mixed models by using smoothing splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61, 381–400. doi:10.1111/1467-9868.00183.

MacNab, Y.C., Dean, C.B., 2002. Spatio-temporal modelling of rates for the construction of disease maps. Statistics in Medicine 21, 347–358.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.P., Guinet, C., 2005. Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in Northwestern Mediterranean Sea, in: Leuangthong, O., Deutsch, C.V. (Eds.), Geostatistics Banff 2004. Springer Netherlands. volume 14 of *Quantitative Geology and Geostatistics*, pp. 777–786. doi:10.1007/978-1-4020-3610-1_81.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.P., Guinet, C., 2006. Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecological Modelling 193, 615–628. doi:10.1016/j.ecolmodel.2005.08.042.

Mugglin, A.S., Carlin, B.P., 1998. Hierarchical modeling in geographic information systems: population interpolation over incompatible zones. Journal of Agricultural, Biological, and Environmental Statistics 3, 111–130. URL: http://www.jstor.org/stable/1400646.

Mugglin, A.S., Carlin, B.P., Gelfand, A.E., 2000. Fully model-based approaches for spatially misaligned data. Journal of the American Statistical Association 95, 877–887. doi:10.2307/2669471.

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554. doi:10.2307/2334389.

Perperoglou, A., Eilers, P.H.C., 2010. Penalized regression with individual deviance effects. Computational Statistics 25, 341–361. doi:10.1007/s00180-009-0180-x.

Pickle, L., Mungiole, M., Jones, G.K., White, A.A., 1999. Exploring spatial patterns of mortality: the new Atlas of United States Mortality. Statistics in Medicine 18, 3211–3220. doi:10.1002/(SICI)1097-0258(19991215)18:233.0.CO;2-Q.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org.

Ramsay, T., 2002. Spline smoothing over difficult regions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 307–319. doi:10.1111/1467-9868.00339.

Rey, G., Jougla, E., Fouillet, A., Hémon, D., 2009. Ecological association between a deprivation index and mortality in France over the period 1997-2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. BMC Public Health 9. doi:10.1186/1471-2458-9-33.

Rodríguez-Álvarez, M.X., Lee, D.-J., Kneib, T., Durbán, M., Eilers, P.H.C., 2015. Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm. Statistics and Computing 25, 941–957. doi:10.1007/s11222-014-9464-2.

Salmond, C.E., Crampton, P., 2012. Development of New Zealand's deprivation index (NZDep) and its uptake as a national policy tool. Canadian Journal of Public Health 103, S7–S11.

Schall, R., 1991. Estimation in generalized linear models with random effects. Biometrika 78, 719–721.

Thompson, R., Baker, R.J., 1981. Composite link functions in generalized linear models. Journal of the Royal Statistical Society: Series C (Applied Statistics) 30, 125–131. URL: http://www.jstor.org/stable/2346381.

Ver Hoef, J.M., 2012. Who invented the delta method?. The American Statistician 66, 124–127. doi:10.1080/00031305.2012.687494.

Waller, L.A., Gotway, C.A., 2004. Applied Spatial Statistics for Public Health Data. John Wiley & Sons, New York.

Wood, S.N., Bravington, M.V., Hedley, S.L., 2008. Soap film smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70, 931–955. doi:10.1111/j.1467-9868.2008.00665.x.

Zhu, L., Carlin, B.P., English, P., Scalf, R., 2000. Hierarchical modelling of spatio-temporally misaligned data: relating traffic density to paediatric asthma hospitalizations. Environmetrics 11, 43–61.