Spatial Modelling of Repeated Events with an Application to Disease Mapping

Debarghya Jana[1] **Niranjan Dey**[2] **Rishiraj Sutar**[3] **Shatrughna Chaurasia**[4]

Roll: 221306 **Roll: 221361** **Roll: 221392** **Roll: 221417**

Supervisor: Dr. Arnab Hazra[5]

A Project Report Submitted in the Requirements of the course MTH643A for the Degree of

MASTER OF SCIENCE

in

STATISTICS

to

DEPARTMENT OF MATHEMATICS AND STATISTICS

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

[1]final year Student, M.Sc. Statistics, IIT Kanpur
[2]final year Student, M.Sc. Statistics, IIT Kanpur
[3]final year Student, M.Sc. Statistics, IIT Kanpur
[4]final year Student, M.Sc. Statistics, IIT Kanpur
[5]Assistant Professor, Department of Mathematics and Statistics, IIT Kanpur

ABSTRACT

Mixed models are frequently employed in the analysis of spatial data, especially in fields such as health sciences and life studies. Spatial random effects are commonly integrated into these models to accommodate the inherent spatial variability present in the data. In the context of spatial count data analysis, Poisson mixed models are often utilized. However, a common assumption in these models is the independence of observations within each area, given the spatial random effects. This assumption might not hold in practical scenarios. For example, when considering multiple asthma visits by a child to physicians within a year, these visits may not be independent observations. To address this limitation, the paper proposes the development of spatial models that account for repeated events. Specifically, compound Poisson mixed models are introduced to simultaneously handle the spatial variation in the data and incorporate the dependencies among repeated events. This innovative approach aims to provide a more realistic representation of the complex nature of the data. The performance of the proposed methodology is assessed through simulation studies, and its practical utility is demonstrated using a real dataset related to children's asthma visits to physicians in the province of Manitoba, Canada. This research contributes to the refinement of spatial modelling techniques, offering a more accurate and nuanced analysis of spatial count data with repeated events.

Keywords: Compound Poisson, Conditional auto-regressive model, Quasi-likelihood, Random effects, Spatial data.

Contents

1. Introduction

The exploration of disease occurrence or mortality patterns across diverse regions has become a focal point, driven by the growing demand for precise disease mapping. The aim of advancing spatial modelling of disease incidence is to adeptly capture variations in authentic disease patterns and discern systematic variability from random noise, which frequently obscures the underlying disease incidence maps. A map depicting the spatial distribution of disease incidence serves as a valuable instrument for recognizing patterns and strategically allocating resources. Disease incidence rates can exhibit substantial variations across geographic areas, and obtaining a dependable estimate of the underlying disease risk frequently entails incorporating information from neighbouring regions. While Poisson regression is a common method for analyzing disease cases, it inherently assumes independence among cases nearby and presupposes that the variance of the response is equal to the mean. However, these assumptions may not always hold, as unmeasured or unknown causal factors for the disease, omitted from the regression model, can contribute to extra-Poisson variation. Furthermore, a certain degree of spatial correlation may be induced in the response, depending on how smoothly the omitted factors vary across the areas. (Clayton and Kaldor, 1987) extended the use of mixed models for geographical data to account for the extra-Poisson variability as well as spatial correlation through the incorporation of spatial random effects in the context of disease mapping. To capture spatial random effects, different forms of conditional autoregressive (CAR) models have been introduced such as the Intrinsic CAR model (Besag, York and Mollié, 1991), the Proper CAR model, and the Leroux CAR (LCAR) model (Leroux, Lei and Breslow, 2000). The LCAR model considers separate parameters for overdispersion and the strength of spatial dependence. Also, the range of spatial dependence parameter is between 0 and 1 for the LCAR model which makes an easy interpretation unlike the proper CAR model. There are many different ways to perform inference in mixed models and in particular generalized linear mixed models (GLMMs). With advances in computational power, one may want to use Markov chain Monte Carlo (MCMC) methods such as Gibbs sampler or Metropolis-Hastings algorithm (Bernardinelli and Montomoli, 1992). The method of penal-

ized quasi-likelihood (PQL) (Breslow and Clayton, 1993) may also be used for inference in the GLMMs. This method was first proposed by and they provided an example of the use of PQL for estimation in mapping studies. Monte Carlo Expectation–Maximization (MCEM) approach (McCulloch, 1997) and Monte Carlo Newton– Raphson (MCNR) algorithm was also used for inference in the GLMMs introduced a frequentist approach, called data cloning (DC), to compute the maximum likelihood estimates (MLE) and their standard errors for general hierarchical models described an approach to compute prediction and prediction interval of random effects in the context of GLMMs. Recently, there has been consideration of applying the DC method to spatial and spatiotemporal Poisson models. As an alternative approach, researchers have explored the use of the generalized estimating equation (GEE) method in the analysis of longitudinal data within the framework of generalized linear models (GLMs). Although GEE was initially developed for longitudinal data with the assumption of independence among subjects or clusters, researchers have extended its application to spatial generalized linear mixed models (GLMMs). This adaptation involves overlooking certain spatial dependencies in the data to apply GEE to spatial contexts. To overcome this issue, Lin considered the quasi-likelihood (QL) approach (Lin and Clayton, 2005) to account for the full spatial covariance structure of the data. (Torabi, 2013) studied spatiotemporal Poisson models with using the QL approach for estimating fixed effect parameters and the GEE to estimate the corresponding variance components. In the all aforementioned models, it is assumed that the number of counts in each area, conditional on the spatial random effects, are independent of each other. This may not be a valid assumption in many situations. For instance, as a motivation of this paper, multiple asthma visits to physicians within a year by a child, called repeated events, are not clearly independent observations. Neglecting the dependency between repeated events can yield misleading results in practice. Repeated events occur in many chronic diseases. Within the scope of repeated events, studied the GLMM framework (Tascheri, Saavedra-Nievas and Roa-Ureta, 2010) with several candidate models including delta lognormal, delta-gamma, quasi-Poisson and compound Poisson. introduced a compound Poisson method to account for repeated events under the frame of disease cluster

detection. Taking into account the spatial models, a spatial Poisson model for the number of claims, while claim size was modelled using a gamma distribution. They integrated both covariates and spatial random effects into the model, employing a Bayesian approach for inference. As far as our current knowledge extends, the exploration of spatial modelling for repeated events is limited. Emphasizing the significance of accounting for repeated events, especially in the context of chronic diseases, and concurrently addressing spatial dependency in disease mapping, can enhance our comprehension of spatial disease trends and identify potential risk factors for preventive measures. In pursuit of this objective, they introduced a spatial discrete compound Poisson model to accommodate repeated events and capture the spatial variation of the outcome. The paper is organized as follows. The spatial compound Poisson model is introduced in section 2. In section 3, we propose to use the QL to estimate the model parameters including fixed effect parameters and variance components of the spatial random effects as well as the corresponding standard errors. We also derive the smoothed disease ratio in the context of the spatial compound Poisson model in section 4. In section 5, the proposed approach is employed to analyze a real dataset of children's asthma visits to physicians in the province of Manitoba, Canada, during 2000–2009. We also evaluate our approach using simulation studies in section 6, and concluding remarks are given in section 7. Technical details and computer codes are deferred to the Appendix in section 8.

2. Methodology

Imagine that our population analysis is segmented into m distinct regions, such as counties, provinces, or municipalities. The data related to the outcomes are presented as counts, representing variables like the number of disease cases in each respective area. Let C_{iw} denote the random variable that represents the count of individuals experiencing exactly w events in area i (where i ranges from 1 to m). Here, w represents the possible number of repetitions, such as the occurrence of asthma visits to physicians. The aggregated variable C_i is then defined as the sum over all possible values of w for the given area i, representing the total number of cases in that specific area. Let F_{ij} denote the number of events for the

jth individual in area i, where j ranges from 1 to C_i (the total number of cases in area i). Consequently, Y_i is defined as the random variable representing the total number of events at area i and is calculated as the sum of the product of the possible number of repetitions w and the count variable C_{iw}. In simpler terms, Y_i is the sum over all individuals in area i of their respective event counts.

2.1. Statistical model and assumptions

Let F_{ij} follow a Poisson distribution with parameter λ_w, and C_i 's given random effects independently follow Poisson distribution with parameter λ_i. Hence, Y_i 's given the random effects have a compound Poisson distribution with mean $\lambda_w\lambda_i$ and variance $\lambda_w\lambda_i\,(1+\lambda_w)$. In particular, we can write

$$\lambda_i = \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right) \tag{2.1}$$

where e_i is the expected number of events (or otherwise) at area i as an offset, $\mathbf{x}_i^\top(1\times p)$ is a vector of covariates at area $i, \boldsymbol{\beta}(p\times 1)$ is a vector of unknown regression coefficients, $\mathbf{z}_i^\top(m\times 1)$ is a known design vector, and $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_m)^\top$ represent spatial random effects. In particular, the LCAR model (Leroux et al., 2000) is used to capture the spatial random effects η. We consider the following general model for the spatial random effects η :

$$\begin{aligned} &\boldsymbol{\eta} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right) \\ &\boldsymbol{\Sigma}_\eta = \sigma_\eta^2\left[\left(1-\lambda_\eta\right)\mathbf{I_m} + \lambda_\eta \mathbf{R}\right]^{-1} \end{aligned} \tag{2.2}$$

where $\mathbf{I_m}$ is the identity matrix of dimension m; $\mathbf{R}$ is a $m\times m$ intrinsic autoregressive matrix with elements $R_{ii} = \aleph_i$ where $\aleph_i$ is the number of areas that are adjacent to area i; if $i \neq j$ then $R_{ij} = -I\{i\sim j\}$ when $I\{i\sim j\}$ is the indicator of whether regions i and j are neighbors; σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence lying in the interval $[0,1]$. This specification yields the independence case if $\lambda_\eta = 0$, and intrinsic autoregression if $\lambda_\eta = 1$.

2.2. Full model:

We can then write our full model in the form of GLMM as:

$$g[E(\boldsymbol{Y} \mid \boldsymbol{\eta})] = \text{ offset } + \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\eta}, \tag{2.3}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top ; g(\cdot) = [\log(\cdot) - \log \lambda_w]$, the offset is the known vector of the logarithm of the e_i; the covariate matrix $\mathbf{x} = \left[\boldsymbol{J}, \left\{\{x_{ij}\}_{i=1}^{m}\right\}_{j=1}^{p}\right]$ corresponds to the fixed effects and has dimension $m \times \{p+1\}$, where $\boldsymbol{J}$ is the $m \times 1$ vector of ones; and the design matrix $\mathbf{z}$ is the identity matrix with dimension $m \times m$.

2.3. Naive Model: Spatial Poisson model

We choose a Poisson regression model (Gschlößl and Czado, 2007) with spatial effects. In particular, we assume for the number of claims $Y_i, i = 1, ., n$, observed at m regions

$$Y_i \sim \text{Poisson}\,(\mu_i)$$

with mean μ_i given by

$$\mu_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta} + log(t_i)\right)$$

where $log(t_i)$ denotes the offset for ith region. The covariate vector for the i-th observation including an intercept is given by $\mathbf{x}_i = (1, x_{i1}, .., x_{ip})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, .., \beta_p)'$ denotes the vector of unknown regression parameters. Spatial dependencies are modelled by introducing a random effect $\gamma_i, i = 1, ., m$ for each region.$\mathbf{z}_i^\top (m \times 1)$ is a known design vector, and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^\top$ represent spatial random effects.

We consider the following general model for the spatial random effects η :

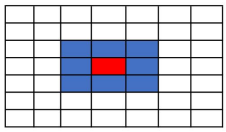
$$\begin{aligned} \boldsymbol{\eta} &\sim N\,(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\ \boldsymbol{\Sigma}_\eta &= \sigma_\eta^2 \left[(1 - \lambda_\eta)\,\mathbf{I_m} + \lambda_\eta \mathbf{R}\right]^{-1} \end{aligned} \tag{2.4}$$

Here, $\mathbf{I_m}$ is the identity matrix of dimension m; $\mathbf{R}$ is a $m \times m$ intrinsic autoregressive matrix

with elements $R_{ii} = \aleph_i$ where $\aleph_i$ is the number of areas that are adjacent to area i; if $i \neq j$ then $R_{ij} = -I\{i \sim j\}$ when $I\{i \sim j\}$ is the indicator of whether regions i and j are neighbors; σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence lying in the interval $[0, 1]$. This specification yields the independence case if $\lambda_\eta = 0$, and intrinsic autoregression if $\lambda_\eta = 1$.

Here $I\{i \sim j\}$ defined as follows, For the red region, $I\{i \sim j\} = \begin{cases} 1 & \text{if } j \in N_i \\ 0 & o/w \end{cases}$

where N_i is the set of blue regions.



3. Quasi-likelihood approach

To estimate the model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$ where $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}, \lambda_w)^\top$ and $\boldsymbol{\theta}_2 = \left(\lambda_\eta, \sigma_\eta^2\right)^\top$ using the QL approach, we first need to find the marginal mean and marginal variance-covariance of $\boldsymbol{Y}$. In particular, to obtain the marginal mean of $Y_i, (i = 1, \ldots, m)$, we can write $\mu_i(\boldsymbol{\theta}) \equiv E(Y_i) = E[E(Y_i \mid \eta)] = \lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) M_\eta(\mathbf{z}_i)$, where $M_\eta(\mathbf{z}_i) = \exp\left(\mathbf{z}_i^\top \boldsymbol{\Sigma}_\eta \mathbf{z}_i / 2\right) = \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)$ and Σ_η^{ii} is the i th diagonal element of $\boldsymbol{\Sigma}_\eta$. To get the marginal variance-covariance of $Y_i, (i = 1, \ldots, m)$, we can write

$$\sigma_{ii}(\boldsymbol{\theta}) \equiv \operatorname{Var}(Y_i) = \mu_i \left\{ \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right) \right] + (1 + \lambda_w) \right\}, \tag{3.1}$$

and,

$$\begin{aligned} \sigma_{ij}(\boldsymbol{\theta}) \equiv \operatorname{cov}(Y_i, Y_j) =& \lambda_w^2 \exp\left[\log e_i + \log e_j + (\mathbf{x}_i + \mathbf{x}_j)^\top \beta\right] \left\{ \exp\left[\frac{1}{2}\left(\Sigma_\eta^{ii} + \Sigma_\eta^{ij}\right)\right] \right. \\ & \left. \times \left[\exp\left(\Sigma_\eta^{ij}\right) - 1\right] \right\} \end{aligned} \tag{3.2}$$

3.1. Quasi-likelihood equations for fixed effects

We define $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_m)^\top$ as the mean vector of the response vector $\boldsymbol{Y}$, and $\mathbf{V}_1(\boldsymbol{\theta})$ as the $m \times m$ variance-covariance matrix of $\boldsymbol{Y}$. We can then write the QL estimating equations for fixed effects as:

$$\mathbf{D}_1^\top \left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right) \mathbf{V}_1^{-1} \left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right) \left[\boldsymbol{Y} - \boldsymbol{\mu} \left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right] = \mathbf{0}$$

where $\mathbf{D}_1(\boldsymbol{\theta}) = \partial \boldsymbol{\mu}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1$ and $\hat{\boldsymbol{\theta}}_2$ is the estimation of $\boldsymbol{\theta}_2$ (see Appendix A.3 for derivation of $\mathbf{D}_1(\boldsymbol{\theta})$). One can use the Newton-Raphson iterative approach to estimate $\boldsymbol{\theta}_1$. To this end, given the value $\hat{\boldsymbol{\theta}}_1^{(k)}$ at the k th iteration, $\hat{\boldsymbol{\theta}}_1^{(k+1)}$ is obtained at the $(k+1)$ th iteration as:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1^{(k+1)} \approx \hat{\boldsymbol{\theta}}_1^{(k)} &+ \left[\mathbf{D}_1^\top \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right) \mathbf{V}_1^{-1} \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right) \mathbf{D}_1 \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \\ &\times \left\{\mathbf{D}_1^\top \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right) \mathbf{V}_1^{-1} \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right) \left[\boldsymbol{Y} - \boldsymbol{\mu} \left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]\right\}. \end{aligned} \tag{3.3}$$

One can then get the variance of $\hat{\boldsymbol{\theta}}_1$ which can be estimated by

$$\widehat{\operatorname{var}\left(\hat{\boldsymbol{\theta}}_1\right)} \approx \left[\mathbf{D}_1^\top(\hat{\boldsymbol{\theta}}) \mathbf{V}_1^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{D}_1(\hat{\boldsymbol{\theta}})\right]^{-1} \tag{3.4}$$

3.2. Quasi-likelihood equations for variance components

In a similar way, we can define $\boldsymbol{S}(\boldsymbol{\theta}) = (S_1, S_2, \ldots, S_m)^\top$ where $S_i = (Y_i - \mu_i)^2, (i = 1, \ldots, m)$; $\boldsymbol{\sigma}(\boldsymbol{\theta}) = (\sigma_{11}, \sigma_{22}, \ldots, \sigma_{mm})^\top$ as the mean vector of $\boldsymbol{S}(\boldsymbol{\theta})$ (see Eqs. (2) and (3)); and $\mathbf{V}_2(\boldsymbol{\theta}) = \operatorname{cov}(\boldsymbol{S}) = \{\mathcal{V}_{ij}(\boldsymbol{\theta})\}_{i,j=1}^m$ as the $m \times m$ variance-covariance matrix of $\boldsymbol{S}(\boldsymbol{\theta})$, (see Appendix A.5 for derivation of $\mathbf{V}_2(\boldsymbol{\theta})$). Hence, the QL estimating equation for variance components can be written as:

$$\mathbf{D}_2^\top \left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) \mathbf{V}_2^{-1} \left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) \left[\boldsymbol{S} \left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) - \boldsymbol{\sigma} \left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right] = \mathbf{0}, \tag{3.5}$$

where $\mathbf{D}_2(\boldsymbol{\theta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_2$ and $\hat{\boldsymbol{\theta}}_1$ is the estimation of $\boldsymbol{\theta}_1$ (see Appendix A.6 for derivation of $\mathbf{D}_2(\boldsymbol{\theta})$). We can then use, for example, the Newton-Raphson iterative to estimate the

variance components $\boldsymbol{\theta}_2$ from the QL estimating equations (6). In particular, given the value $\hat{\boldsymbol{\theta}}_2^{(k)}$ at the k th iteration, $\hat{\boldsymbol{\theta}}_2^{(k+1)}$ is obtained at the $(k+1)$ th iteration as:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_2^{(k+1)} \approx \hat{\boldsymbol{\theta}}_2^{(k)} + \left[\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]^{-1} \\ \times\left\{\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\left[\boldsymbol{S}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]\right\}\end{aligned} \tag{3.6}$$

The variance of $\hat{\boldsymbol{\theta}}_2$ can be also estimated by (see Appendix A.7 for detail):

$$\widehat{\text{var}\left(\hat{\boldsymbol{\theta}}_2\right)} \approx \left\{\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}})\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}})\mathbf{D}_2(\hat{\boldsymbol{\theta}})\right\}^{-1} \tag{3.7}$$

3.3. Complete algorithm to estimate the model parameters

Algorithm 1 Complete algorithm to estimate the model parameters

1: Choose initial values $\boldsymbol{\theta^0} = (\boldsymbol{\theta_1^0}, \boldsymbol{\theta_2^0})$ as mentioned in the paper.

2: Calculate $\hat{\boldsymbol{\theta}}_1^{(k+1)}$ from,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1^{(k+1)} \approx \hat{\boldsymbol{\theta}}_1^{(k)} + \left[\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \\ \times\left\{\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]\right\}.\end{aligned} \tag{3.8}$$

3: Calculate $\hat{\boldsymbol{\theta}}_2^{(k+1)}$ from,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_2^{(k+1)} \approx \hat{\boldsymbol{\theta}}_2^{(k)} + \left[\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]^{-1} \\ \times\left\{\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\left[\boldsymbol{S}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]\right\}\end{aligned} \tag{3.9}$$

4: Stop when $\|\boldsymbol{\theta^{(k+1)}} - \boldsymbol{\theta^{(k)}}\| < \epsilon$.

4. Disease ratio

One of our main interests in spatial statistics and in particular in disease mapping is to map the prediction of the disease ratio (or rate). The smoothed disease ratio (SDR) for each area can be written as:

$$SDR_i = \frac{\hat{\lambda}_w\hat{\lambda}_i}{e_i} = \hat{\lambda}_w \exp\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \hat{\eta}_i\right) \tag{4.1}$$

where $\hat{\eta}_i$ is the predicted spatial random effect $\left(\mathbf{z}_i^\top \boldsymbol{\eta} = \eta_i\right)$ of disease at area i given by

$$\hat{\eta}_i = \widehat{E\left(\eta_i \mid Y_i\right)} = E\left(\eta_i \mid Y_i\right)|_{\theta=\hat{\theta}} \tag{4.2}$$

which is the posterior mean (best predictor) of η_i, given the areal disease count Y_i evaluated at $\hat{\boldsymbol{\theta}}$. Using the Bayes Theorem and our model assumption, we get

$$\begin{aligned}
E\left(\eta_i \mid Y_i\right) =& \frac{\int_{-\infty}^{\infty} \eta_i P_{Y_i|\eta_i}\left(Y_i \mid \eta_i\right) f_{\eta_i}\left(\eta_i\right) d\eta_i}{\int_{-\infty}^{\infty} P_{Y_i|\eta_i}\left(Y_i \mid \eta_i\right) f_{\eta_i}\left(\eta_i\right) d\eta_i}, \\
E_i^r \equiv& \int_{-\infty}^{\infty} \eta_i^r P_{Y_i|\eta_i}\left(Y_i \mid \eta_i\right) f_{\eta_i}\left(\eta_i\right) d\eta_i \\
\approx& \int_{-\infty}^{\infty} {\eta_i}^r \frac{e^{-\lambda_i} \lambda_w^{Y_i}}{Y_i!} \sum_c \frac{e^{-c\lambda_w} \lambda_i^c c^{Y_i}}{c!} \times \frac{1}{\sqrt{2\pi\Sigma_\eta^{ii}}} e^{-\frac{{\eta_i}^2}{2\Sigma_\eta^{ii}}} d\eta_i \\
=& \frac{\lambda_w^{Y_i}}{\sqrt{2\pi\Sigma_\eta^{ii}} Y_i!} \sum_c \frac{K_i^c c^{Y_i} e^{-c\lambda_w}}{c!} \int_{-\infty}^{\infty} \exp\left\{ r \log \eta_i + c\eta_i - K_i e^{\eta_i} - \frac{{\eta_i}^2}{2\Sigma_\eta^{ii}} \right\} d\eta_i,
\end{aligned} \tag{4.3}$$

where $r = 0, 1$ and $K_i = e_i e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$. Hence, by Laplace approximation (Tascheri, Saavedra-Nievas and Roa-Ureta, 2010), we can estimate $E\left(\eta_i \mid Y_i\right)$ as:

$$\begin{aligned}
\hat{\eta}_i = \widehat{E\left(\eta_i \mid Y_i\right)} =& \left. \frac{E_i^1}{E_i^0} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\
\approx& \frac{\sum_c \frac{\hat{K}_i^c c^{Y_i} e^{-c\hat{\lambda}_w}}{c!} \sqrt{-\frac{1}{\hat{L}''_{1,i}\left(\eta^*_{1,i}\right)}} e^{\hat{L}_{1,i}\left(\eta^*_{1,i}\right)}}{\sum_c \frac{\hat{K}_i^c c^{Y_i} e^{-c\hat{\lambda}_w}}{c!} \sqrt{-\frac{1}{\hat{L}''_{0,i}\left(\eta^*_{0,i}\right)}} e^{\hat{L}_{0,i}\left(\eta^*_{0,i}\right)}},
\end{aligned} \tag{4.4}$$

where

$$L_{r,i}\left(\eta_i\right) = \left(r \log \eta_i + c\eta_i - K_i e^{\eta_i} - \frac{\eta_i^2}{2\Sigma_\eta^{ii}} \right)$$

and

$$L''_{r,i}\left(\eta_i\right) = -\left(\frac{r}{\eta_i^2} + K_i e^{\eta_i} + \frac{1}{\Sigma_\eta^{ii}} \right)$$

where $\eta^*_{r,i}$ is the mode of $L_{r,i}\left(\eta_i\right)$ and can be obtained numerically through the R software (Team, 2010) built-in function optim.

5. Application

We use a dataset of children (age < 18 years) asthma visits to physicians in the Canadian province of Manitoba during 2000-2009 to apply our proposed approach. The population of Manitoba was stable during the study period from 1.15 million in 2000 to 1.22 million in 2009. The province consisted of five regional health authorities that were responsible for the delivery of health care services. These five regions were further subdivided into 67 Regional Health Authorities Districts (RHAD). The RHAD are the geographic units (areas) used in our model; all data are linked to these geographic boundaries. The number of children's asthma visits totaled 694,484 over the study period with mean and median number of yearly cases per area of 1037 and 493 (range 36 to 5718), respectively. The average population of children with the age under 18 years was about 302,949. The average child population sizes varied from 287 to 21,966, with mean and median numbers of 4522 and 2678, respectively. We provided a histogram of the number of visits for asthma to physicians by individuals (re-admissions) during the study period.

	w	C_w
1	1	205657
2	2	76478
3	3	30164
4	4	17254
5	5	8926
6	6	5493
7	7	3358
8	8	2209
9	9	1412
10	10	1038
11	11	627
12	12	457
13	13	348
14	14	249
15	15	197
16	16	144
17	17	99
18	18	59
19	19	46
20	20	50
21	21	25
22	22	28
23	23	19
24	24	16
25	25	12
26	26	13
27	27	4
28	28	10
29	29	3
30	30	7
31	31	2
32	32	5
33	34	3
34	35	2
35	36	2
36	37	1
37	39	3
38	40	2
39	41	1
40	43	1
41	47	1
42	56	1
43	68	1

FIG 1. *w is the number of repeated events and C_w is the number of visits*

It is clear that we have range 1 to 68 number of visits as repeated events from the same individuals rather than from different children. It shows that assuming the all visits in each area are independent of each other is not a valid assumption. Model parameter estimates (EST) and corresponding standard errors (SE) using spatial compound Poisson model for childhood asthma visits in Manitoba, Canada, during 2000-2009. Note that the expected number of asthma cases e_i is adjusted by sex and year. In particular, we have $e_i = \sum_{j=1}^{2} \sum_{t=1}^{10} n_{ijt} \frac{y_{jt}}{n_{jt}}$ where n_{ijt} is the population at risk for the i th area, sex j, and year t; n_{jt} is the population at risk for the sex j and year t which is given by $n_{jt} = \sum_{i=1}^{67} n_{ijt}$, and similarly, y_{jt} is the number of children asthma visits for the sex j and year t. The model parameter estimates and corresponding standard errors for the proposed model. Note that we estimate and report $\lambda_\eta = 0.5$, $\sigma_\eta = 2$ rather than σ_η^2 for simplicity in computation (Leroux, Lei and Breslow, 2000). It seems that all the model parameters have significant contributions to the model. In particular, it appears that the estimation of $\lambda_w = 5.5$ is significantly larger than 1 which is the basic condition for a naive model which ignores the dependence of observations in each area given the spatial random effects. The spatial random effect parameter estimates also show the significant spatial dependency for our asthma data. The importance of considering the dependency among observations in each area, given the spatial random effects, is more explored in the simulation study.

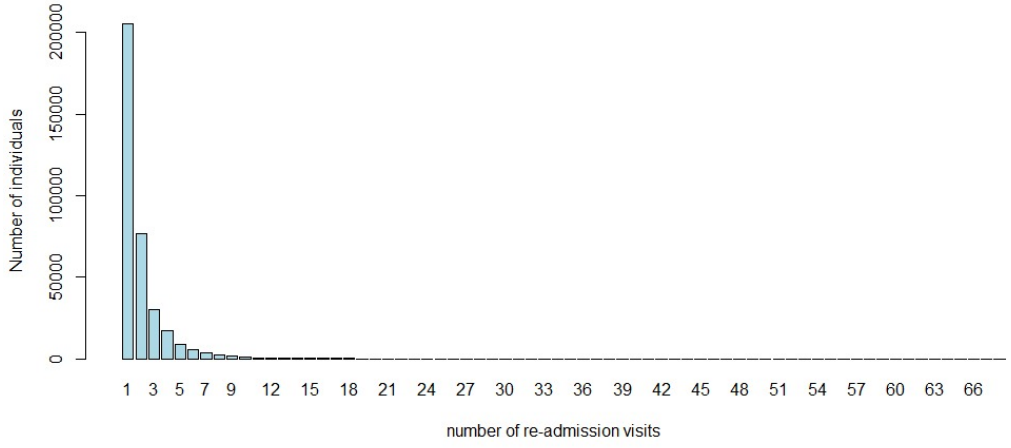


FIG 2. *Barplot of Total number of asthma visits in i^{th} location (Y_i)*

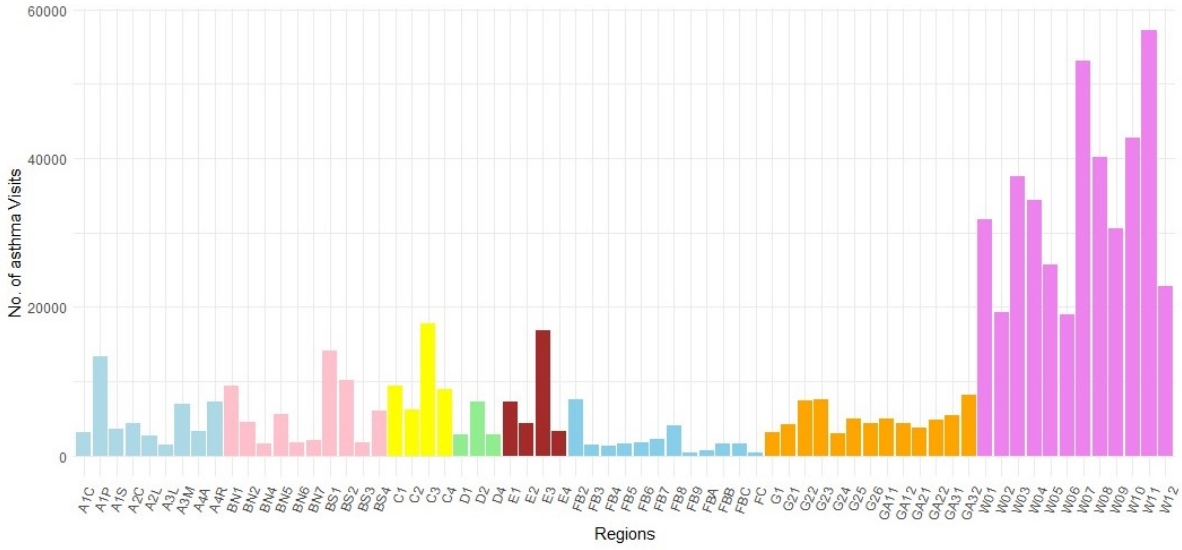


FIG 3. *Barplot of Total number of asthma visits in i^{th} location (Y_i)*

The dataset consists of 11825 rows and 5 columns where the first column indicates the spatial region, the second column indicates the year, third column indicates sex. The fourth column indicates the number of repetitions or w and the last column indicates the number of individuals with 'w' number of repetitions. childhood asthma visits compared to the rest of the population.

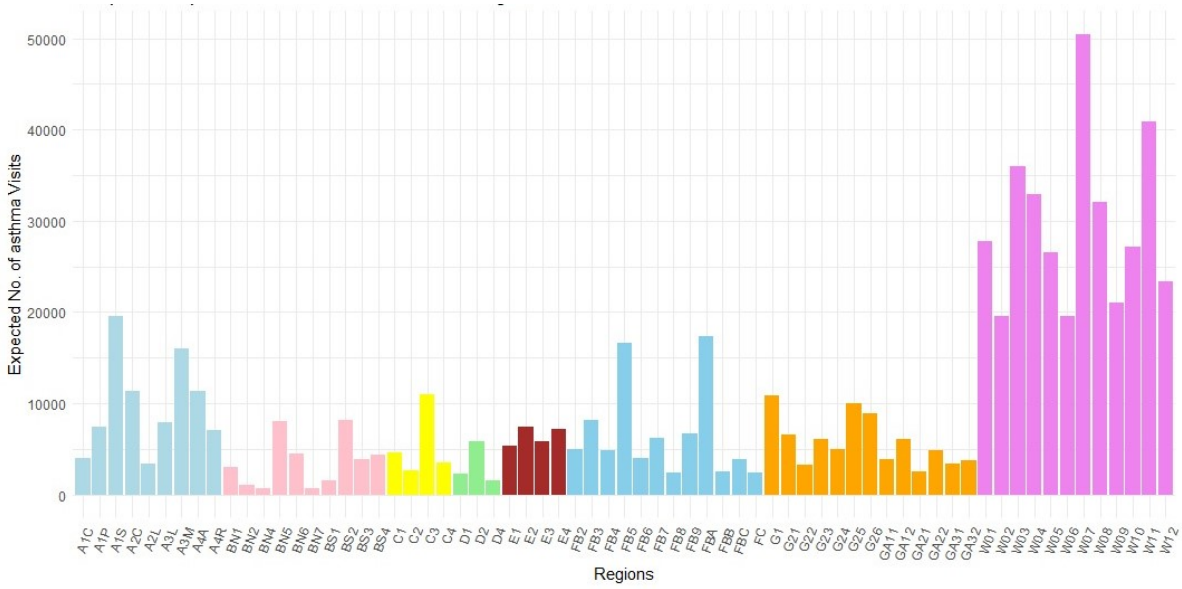


FIG 4. *Barplot of expected number of asthma visits in i^{th} location (e_i)*

These findings may represent real increases or different distributions of important covariates that are unmeasured and unadjusted for in our modelling. We also provide a map of the prediction of spatial random effects. The spatial component resembles the map of disease ratio (Figure) which clearly shows the dominance of a strong underlying spatial structure in our data.

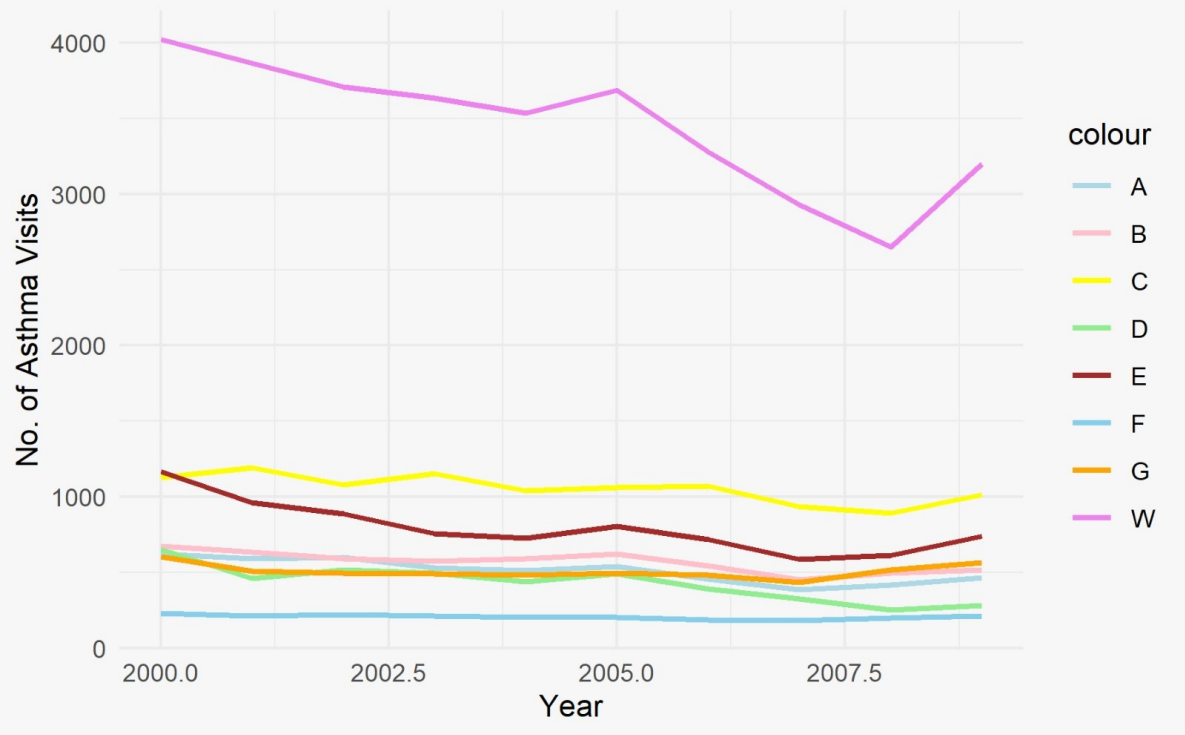


FIG 5. *Line diagram of total no. of asthma visits for different years*

6. Simulation study

Additionally, we carried out simulation studies to assess and contrast the effectiveness of our proposed approach with a naive approach that neglects the interdependence of observations within each area, considering the influence of spatial random effects. We evaluated the bias and mean squared error (MSE) of the model parameter estimates for both the proposed model (spatial compound Poisson model) and the naive model (spatial Poisson model) specifically in the context of an irregular grid. To have a better understanding of our proposed model performance, both irregular and regular neighbourhood structures with

different parameter values are examined. For all the settings, 12,000 datasets are generated from both proposed and naive models.

Parameter	True value	Proposed model		Naive model	
		Bias	MSE	Bias	MSE
β_0	-2.5	-0.46182	1.03920	1.79171	3.21471
λ_w	5.5	-0.07619	0.03425	-	-
λ_η	0.5	-0.04606	0.00215	0.25829	0.06701
σ_η	2	-0.12049	0.01403	-0.48105	0.23171

TABLE 1

Bias and mean squared error (MSE) of the model parameter estimates for the proposed model (spatial compound Poisson model) and naive model (spatial Poisson model) in the case of an irregular grid.

The NewtonRaphson algorithm to estimate the model parameters is updated with a maximum of 500 iterations and at each step the convergence is checked using a tolerance of 0.009. The proper boundary values are also set for the estimates outside of their natural ranges ($\sigma_\eta \geq 0, 0 \leq \lambda_\eta \leq 1$).

Here we use the spatial layout of our Manitoba RHADs, which is an irregular neighborhood structure. Data are generated from the model (2.3) with the true parameters similar to the model parameter estimates reported in Table 1. In particular, the neighbourhood structure and the expected counts are exactly as for the asthma visit dataset. Table 2 presents the bias and mean square error (MSE) of the model parameter estimates for the spatial compound Poisson and naive (spatial Poisson model) models.

We also provide boxplots of model parameter estimates for both proposed and naive models to have a better understanding of the performance of these models. It appears that biases and MSEs of the model parameter estimates for our proposed model are considerably smaller than the corresponding values from the naive model. It is also clear that the variability of estimated parameters in the case of the proposed model is smaller than the naive model, particularly for spatial random effect parameters. It is worth mentioning that the naive model also has very low convergency (about 500 out of 12,000) in comparison with the proposed model mainly due to the negative values of the estimates of the variance of random effects. Hence, to have a fair comparison, only 500 simulation runs of the proposed model are considered.

With the same setting as used in Table 1, we set the parameter values similar to the values obtained in the data application ($\beta_0 = -2.5, \lambda_\eta = 0.5$, and $\sigma_\eta = 2$) with four different values for λ_w as $1, 5, 10$, and 15, in order to evaluate the efficiency of the proposed model with the naive model while the number of events per individual increases. Table 3 shows the biases and MSEs of the model parameter estimates for the both proposed and naive models. As expected, the MSEs of model parameter estimates tend to increase with increasing the number of events per individual (λ_w). However, the biases and MSEs of model parameter estimates in the case of the proposed model are consistently smaller than the corresponding values in the case of the naive model.

Model	True parameter λ_w	β_0		λ_η		σ_η		λ_w	
		− 2.5		0.5		2			
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Proposed	1	-0.20706	0.14723	-0.04939	0.00227	-0.06007	0.00447	-0.13402	0.17224
Naive		-0.56907	1.30002	-0.06772	0.00431	-0.12963	0.01746	-	-
Proposed	5	−0.47316	0.90627	−0.06103	0.00323	−0.12427	0.01503	−0.08141	0.03569
Naive		1.71729	2.95530	0.25042	0.06247	−0.50304	0.25313	-	-
Proposed	10	−0.50941	1.00219	−0.04231	0.00226	−0.14123	0.02003	−0.04832	0.01007
Naive		2.00226	4.01172	0.26429	0.07003	−0.58404	0.34103	-	-
Proposed	15	−0.46824	1.20953	−0.02623	0.00105	−0.17126	0.02918	−0.03007	0.00502
Naive		2.00431	4.01572	0.25424	0.06405	−0.57447	0.32905	-	-

TABLE 2

Bias and mean squared error (MSE) of the model parameter estimates for the proposed model (spatial compound Poisson model) and naive model (spatial Poisson model) for different values of λ_w in the case of irregular grid.

case of $\lambda_w = 1$, we expect to have an equal or better result in terms of estimated parameters for the naive model compared to the proposed model as the naive model has naturally been designed for this case ($\lambda_w = 1$) which leads to smaller variation compared to the proposed model.

As it is also shown in Figure 6, the variability of estimated parameters in the case of the proposed model is noticeably smaller than the naive model in particularly for spatial random effect parameters.

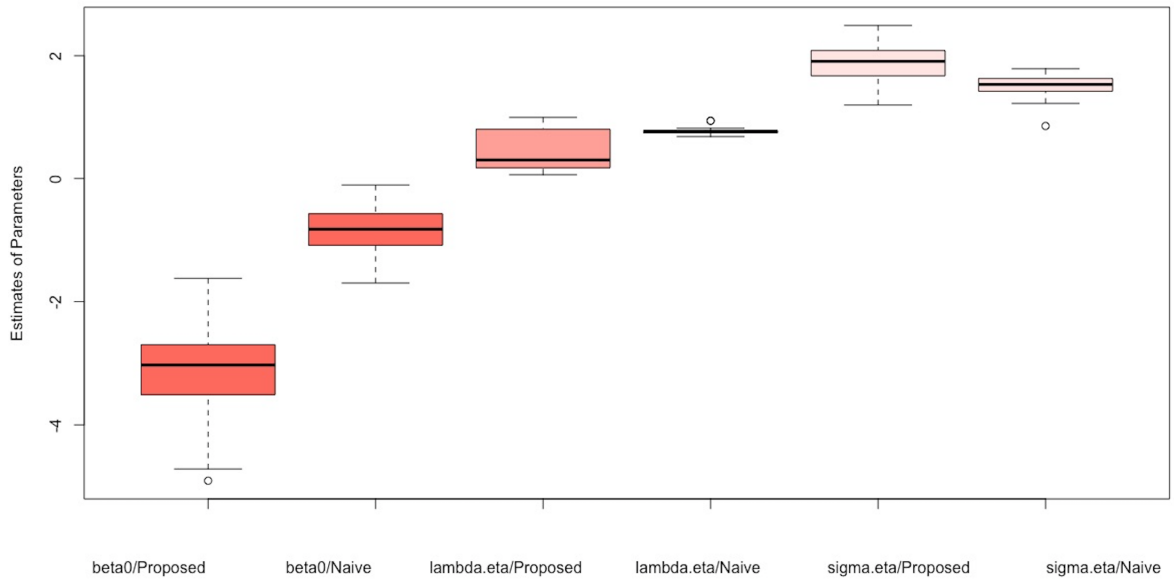


FIG 6. *Box plots of the model parameter estimates for the proposed and naive models in the case of irregular grid*

7. Conclusion

In many applications, there are count data, for example, in an area, conditional on spatial random effects, that are not independent of each other. For instance, in our childhood asthma visit data application, we had on average almost 5 visits (range 1 to 68) as repeated events from the same individuals rather than different children. In this paper, we have proposed a spatial compound Poisson model to account for the repeated events as well as the spatial variation of the data. In particular, the model accommodated a Leroux conditional auto-regressive model for the spatial random effects. We have also proposed to use the quasi-likelihood approach to estimate the model parameters with the corresponding standard errors. The derivation of the smoothed disease ratio over space was also provided.

We adjusted our expected number of children's asthma visits to physicians by important factors of sex and year. We did not have access to other covariates related to asthma, however, our proposed model could accommodate covariates directly into the model. Our simulation studies also clearly showed the outperformance of our proposed model (spatial compound Poisson model) compared to the naive model (spatial Poisson model), which ignores the

repeated events in the model, in terms of bias and mean squared error of model parameter estimates.

Our proposed spatial compound Poisson method is very general in the context of spatial statistics. Our approach opens a new direction for spatial data with repeated events. In this paper, we used spatial count data with repeated events, however, in some applications, we may have spatiotemporal count data with repeated events. For instance, in our application, we may have access to children's asthma visits to physicians' data over space and each time visit for each child. Another interesting contribution will be to extend our proposed approach for binary data which has many applications (Torabi, 2013). These are some of the topics for future study.

8. Appendix

This appendix provides derivation of elements of the QL estimating equations needed to estimate the model parameters and corresponding standard errors.

8.1. Appendix 1

We will derive the link function as $\{\log(\cdot) - \log \lambda_w\}$.We can write

$$\log(\lambda_i) = \text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta} \tag{8.1}$$

then $\lambda_i = \exp(\text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta})$, Also, assuming $E(Y_i \mid \boldsymbol{\eta}) = \lambda_w \lambda_i$, we can write

$$E(Y_i \mid \boldsymbol{\eta}) = \lambda_w \exp\big(\text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\big) \tag{8.2}$$

8.2. Appendix 2

We will derive the $\mathbf{V}_1(\boldsymbol{\theta})$.To obtain $V_1(\boldsymbol{\theta})$, we need to calculate the components of $\text{cov}(\boldsymbol{Y})$. To that end, we need to calculate $\sigma_{ii} = \text{var}\,(Y_i)$ and $\sigma_{ij} = \text{cov}\,(Y_i, Y_j)$ as

$$\begin{aligned}
\sigma_{ii}(\boldsymbol{\theta}) &\equiv \text{var}\,(Y_i) = \text{var}\,[E\,(Y_i \mid \boldsymbol{\eta})] + E\,[\text{var}\,(Y_i \mid \boldsymbol{\eta})] \\
&= \text{var}\left[\lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right)\right] + E\left[E\,(Y_i \mid \boldsymbol{\eta}) + \lambda_w E\,(Y_i \mid \boldsymbol{\eta})\right] \\
&= \mu_i \left\{ \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \left[\lambda_w \frac{M_\eta\,(2\mathbf{z}_i)}{M_\eta\,(\mathbf{z}_i)} - \lambda_w \frac{M_\eta\,(\mathbf{z}_i)^2}{M_\eta\,(\mathbf{z}_i)}\right] + (1 + \lambda_w) \right\} \\
&= \mu_i \left\{ \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + (1 + \lambda_w) \right\},
\end{aligned} \tag{8.3}$$

note that $E\,(Y_i \mid \boldsymbol{\eta}) = \lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right) = \lambda_w \lambda_i$ and $\text{var}\,(Y_i \mid \boldsymbol{\eta}) = \lambda_w \lambda_i\,(1 + \lambda_w) = E\,(Y_i \mid \boldsymbol{\eta}) + \lambda_w E\,(Y_i \mid \boldsymbol{\eta})$. Similarly, we can write $\text{cov}\,(Y_i, Y_j)\,, (i \neq j)$, as

$$\begin{aligned}
\sigma_{ij}(\boldsymbol{\theta}) &\equiv \text{cov}\,(Y_i, Y_j) = \text{cov}\,[E\,(Y_i \mid \boldsymbol{\eta})\,, E\,(Y_j \mid \boldsymbol{\eta})] + E\,[\text{cov}\,(Y_i, Y_j \mid \boldsymbol{\eta})] \\
&= \lambda_w^2\,\text{cov}\,(\lambda_i, \lambda_j) + E[0] \\
&= \lambda_w^2\,\text{cov}\left[\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right), \exp\left(\log e_j + \mathbf{x}_j^\top \boldsymbol{\beta} + \mathbf{z}_j^\top \boldsymbol{\eta}\right)\right] \\
&= \lambda_w^2 \exp\left[\log e_i + \log e_j + (\mathbf{x}_i + \mathbf{x}_j)^\top \boldsymbol{\beta}\right] [M_\eta\,(\mathbf{z}_i + \mathbf{z}_j) \\
&- M_\eta\,(\mathbf{z}_i)\,M_\eta\,(\mathbf{z}_j)] \\
&= \lambda_w^2 \exp\left[\log e_i + \log e_j + (\mathbf{x}_i + \mathbf{x}_j)^\top \beta\right] \left\{ \exp\left[\frac{1}{2}\left(\Sigma_\eta^{ii} + \Sigma_\eta^{ij}\right)\right] \right. \\
&\left. \times \left[\exp\left(\Sigma_\eta^{ij}\right) - 1\right] \right\}.
\end{aligned} \tag{8.4}$$

We then have

$$g\,[E\,(Y_i \mid \boldsymbol{\eta})] = \text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}, \tag{8.5}$$

where $g(\cdot) = \{\log(\cdot) - \log \lambda_w\}$.

8.3. Appendix 3

We will derive the $\mathbf{D}_1(\boldsymbol{\theta})$ here.To obtain $\mathbf{D}_1(\boldsymbol{\theta}) = \{\mathbf{D}_{1i}(\boldsymbol{\theta})\}_{i=1}^{m}$, we need to calculate the components of $\mathbf{D}_{1i}(\boldsymbol{\theta})$. Consequently, it is enough to derive $\partial\mu_i(\boldsymbol{\theta})/\partial\boldsymbol{\beta}$ and $\partial\mu_i(\boldsymbol{\theta})/\partial\lambda_w, (i = 1, \ldots, m)$. To that end, we can write

$$\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} = \mathbf{x}_i^{\top}\left[\lambda_w \exp\left(\log e_i + \mathbf{x}_i^{\top}\boldsymbol{\beta}\right) M_{\eta}\left(\mathbf{z}_i\right)\right] = \mathbf{x}_i^{\top}\mu_i, \tag{8.6}$$

and

$$\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\lambda_w} = \exp\left(\log e_i + \mathbf{x}_i^{\top}\boldsymbol{\beta}\right) M_{\eta}\left(\mathbf{z}_i\right) = \mu_i/\lambda_w \tag{8.7}$$

8.4. Appendix 4

We will derive the var $\left(\hat{\boldsymbol{\theta}}_1\right)$.By applying Taylor expansion and we can write

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1 \approx \boldsymbol{\theta}_1 &+ \left[\mathbf{D}_1^{\top}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \\ &\times\left\{\mathbf{D}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\left[\mathbf{Y} - \boldsymbol{\mu}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right]\right\}\end{aligned} \tag{8.8}$$

Moreover, since

$$\begin{aligned}&\text{var}\left\{\mathbf{D}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\left[\mathbf{Y} - \boldsymbol{\mu}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right]\right\} \\ &= \left[\mathbf{D}_1^{\top}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right],\end{aligned} \tag{8.9}$$

we then have

$$\text{var}\left(\hat{\boldsymbol{\theta}}_1\right) \approx \left[\mathbf{D}_1^{\top}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \tag{8.10}$$

8.5. Appendix 5

We will derive the $\mathbf{V}_2(\boldsymbol{\theta})$.To obtain $V_2(\boldsymbol{\theta})$, we need to calculate the components of cov($\boldsymbol{S}$). To that end, we need to calculate $\mathcal{V}_{ii} = \text{var}\,(S_i)$ and $\mathcal{V}_{ij} = \text{cov}\,(S_i, S_j)$ as follows.

$$\begin{aligned}\mathcal{V}_{ii} = \text{var}\,(S_i) &= E\,(Y_i - \mu_i)^4 - E^2\,(Y_i - \mu_i)^2 \\ &= E\,(Y_i - \mu_i)^4 - {\sigma_{ii}}^2.\end{aligned} \tag{8.11}$$

We now need to calculate $E\,(Y_i - \mu_i)^4$, assuming $E\,(Y_i \mid \eta) = \mu_{i\eta}$, as:

$$E\,(Y_i - \mu_i)^4 = E\left(Y_i^4\right) - 4\mu_i E\left(Y_i^3\right) + 6\mu_i^2 E\left(Y_i^2\right) - 3\mu_i^4 \tag{8.12}$$

where

$$\begin{aligned}E\left(Y_i^4\right) &= E\left[E\left(Y_i^4 \mid \eta\right)\right] = \left(1 + 7\lambda_w + 6\lambda_w^2 + \lambda_w^3\right) E\left(\mu_{i\eta}\right) + \left(7 + 18\lambda_w + 7\lambda_w^2\right) E\left({\mu_{i\eta}}^2\right) \\ &\quad + 6\left(1 + \lambda_w\right) E\left({\mu_{i\eta}}^3\right) + E\left({\mu_{i\eta}}^4\right), \\ E\left(Y_i^3\right) &= E\left[E\left(Y_i^3 \mid \eta\right)\right] = \left(1 + 3\lambda_w + \lambda_w^2\right) E\left(\mu_{i\eta}\right) + 3\left(1 + \lambda_w\right) E\left({\mu_{i\eta}}^2\right) + E\left({\mu_{i\eta}}^3\right),\end{aligned} \tag{8.13}$$

and

$$E\left(Y_i^2\right) = E\left[E\left(Y_i^2 \mid \eta\right)\right] = \left(1 + \lambda_w\right) E\left(\mu_{i\eta}\right) + E\left({\mu_{i\eta}}^2\right) \tag{8.14}$$

Hence,

$$\begin{aligned}E\,(Y_i - \mu_i)^4 &= \mu_i\left(1 + 7\lambda_w + 6\lambda_w^2 + \lambda_w^3\right) - 4\mu_i^2\left(1 + 3\lambda_w + \lambda_w^2\right) + 6\mu_i^3\left(1 + \lambda_w\right) - 3\mu_i^4 \\ &\quad + \left[\left(7 + 18\lambda_w + 7\lambda_w^2\right) - 12\mu_i\left(1 + \lambda_w\right) + 6\mu_i^2\right] E\left({\mu_{i\eta}}^2\right) \\ &\quad + 2\left[3\left(1 + \lambda_w\right) - 2\mu_i\right] E\left({\mu_{i\eta}}^3\right) + E\left({\mu_{i\eta}}^4\right)\end{aligned} \tag{8.15}$$

where

$$\begin{aligned} E\left(\mu_{i\eta}\right) &= \mu_i = \lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) M_\eta\left(\mathbf{z}_i\right), \\ E\left({\mu_{i\eta}}^2\right) &= \lambda_w^2 \exp\left[2\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\right] M_\eta\left(2\mathbf{z}_i\right), \\ E\left({\mu_{i\eta}}^3\right) &= \lambda_w^3 \exp\left[3\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\right] M_\eta\left(3\mathbf{z}_i\right), \\ E\left({\mu_{i\eta}}^4\right) &= \lambda_w^4 \exp\left[4\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\right] M_\eta\left(4\mathbf{z}_i\right). \end{aligned} \tag{8.16}$$

Finally, we can write

$$\begin{aligned} \mathcal{V}_{ij} =& \operatorname{cov}\left(s_i, s_j\right) = E\left[\left(Y_i - \mu_i\right)^2 \left(Y_j - \mu_j\right)^2\right] - \sigma_{ii}\sigma_{jj} \\ &= E\left\{E\left[\left(Y_i - \mu_i\right)^2 \left(Y_j - \mu_j\right)^2 \mid \eta\right]\right\} - \sigma_{ii}\sigma_{jj} \\ &= E\left\{E\left[\left(Y_i - \mu_i\right)^2 \mid \eta\right] E\left[\left(Y_j - \mu_j\right)^2 \mid \eta\right]\right\} - \sigma_{ii}\sigma_{jj} \\ &= E\left[\operatorname{var}\left(Y_i \mid \eta\right) \operatorname{var}\left(Y_j \mid \eta\right)\right] - \sigma_{ii}\sigma_{jj} \\ &= E\left\{\left[\left(1 + \lambda_w\right)\mu_{i\eta}\right]\left[\left(1 + \lambda_w\right)\mu_{j\eta}\right]\right\} - \sigma_{ii}\sigma_{jj} \\ &= \left(1 + \lambda_w\right)^2 E\left(\mu_{i\eta}\mu_{j\eta}\right) - \sigma_{ii}\sigma_{jj}, \end{aligned} \tag{8.17}$$

where

$$\begin{aligned} E\left(\mu_{i\eta}\mu_{j\eta}\right) &= E\left[\lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right) \lambda_w \exp\left(\log e_j + \mathbf{x}_j^\top \boldsymbol{\beta} + \mathbf{z}_j^\top \boldsymbol{\eta}\right)\right] \\ &= \lambda_w^2 \exp\left[\log e_i + \log e_j + \left(\mathbf{x}_i + \mathbf{x}_j\right)^\top \boldsymbol{\beta}\right] M_\eta\left(\mathbf{z}_i + \mathbf{z}_j\right) \end{aligned} \tag{8.18}$$

8.6. Appendix 6

To find $\mathbf{D}_2(\boldsymbol{\theta}) = \{D_{2i}(\boldsymbol{\theta})\}_{i=1}^m$, we need to obtain the components of $\mathbf{D}_{2i}(\boldsymbol{\theta})$. Consequently, it is enough to calculate $\partial\sigma_{ii}(\boldsymbol{\theta})/\partial\lambda_\eta, \partial\sigma_{ii}(\boldsymbol{\theta})/\partial\sigma_\eta^2, (i = 1, \ldots, m)$. To that end, we can write

$$\begin{aligned} &\frac{\partial\sigma_{ii}(\boldsymbol{\theta})}{\partial\lambda_\eta} = \\ &\frac{1}{2}\mu_i K_{ii}\left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + \left(1 + \lambda_w\right)\right\} \\ &+ \frac{1}{2}\mu_i K_{ii}\left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[3\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right]\right\} \\ &= \frac{1}{2}\mu_i K_{ii}\left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[4\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - 2\exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + \left(1 + \lambda_w\right)\right\} \end{aligned} \tag{8.19}$$

where $K_{ii} = \left(\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \lambda_\eta}\right)^{(ii)}$ which is the i th diagonal element of $\sigma_\eta^{-2}\boldsymbol{\Sigma}_\eta \left[\mathbf{I_m} - \mathbf{R}\right] \boldsymbol{\Sigma}_\eta$. We can also write

$$\begin{aligned}
&\frac{\partial \sigma_{ii}(\boldsymbol{\theta})}{\partial \sigma_\eta^2} = \\
&\frac{1}{2}\mu_i \sigma_\eta^{-2}\Sigma_\eta^{ii}\left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\lambda_w\left[\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + (1+\lambda_w)\right\} \\
&+ \frac{1}{2}\mu_i \sigma_\eta^{-2}\Sigma_\eta^{ii}\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\lambda_w\left[3\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] \\
&= \frac{1}{2}\mu_i \sigma_\eta^{-2}\Sigma_\eta^{ii}\left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right)\lambda_w\left[4\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - 2\exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + (1+\lambda_w)\right\}.
\end{aligned} \tag{8.20}$$

8.7. Appendix 7

We will derive the var $\left(\hat{\boldsymbol{\theta}}_2\right)$.By applying Taylor expansion and using Eq. (6), we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_2 \approx \boldsymbol{\theta}_2 &+ \left[\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right]^{-1} \\
&\times\left\{\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\left[\boldsymbol{S}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right]\right\}.
\end{aligned} \tag{8.21}$$

Moreover, since

$$\begin{aligned}
&\text{var}\left\{\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)\left[\mathbf{S}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right]\right\} \\
&= \left[\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right],
\end{aligned} \tag{8.22}$$

we can then write

$$\text{var}\left(\hat{\boldsymbol{\theta}}_2\right) \approx \left[\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right]^{-1} \tag{8.23}$$

References

Bernardinelli, L. and Montomoli, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in medicine* **11** 983–1007.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics* **43** 1–20.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association* **88** 9–25.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 671–681.

Gschlössl, S. and Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal* **2007** 202–225.

Leroux, B. G., Lei, X. and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials* 179–191. Springer.

Lin, P.-S. and Clayton, M. K. (2005). Analysis of binary spatial data by quasi-likelihood estimating equations.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association* **92** 162–170.

Tascheri, R., Saavedra-Nievas, J. and Roa-Ureta, R. (2010). Statistical models to standardize catch rates in the multi-species trawl fishery for Patagonian grenadier (Macruronus magellanicus) off Southern Chile. *Fisheries Research* **105** 200–214.

Team, R. D. C. (2010). R: A language and environment for statistical computing. *(No Title)*.

Torabi, M. (2013). Spatio–temporal modeling for disease mapping using CAR and B-spline smoothing. *Environmetrics* **24** 180–188.