Spatial modeling of repeated events with an application to disease mapping

Shabnam Balamchi[a], Mahmoud Torabi[a,b,*]

[a] *Department of Statistics, University of Manitoba, MB, R3T 2N2, Canada*
[b] *Department of Community Health Sciences, University of Manitoba, MB, R3E 0W3, Canada*

Abstract

Mixed models are commonly used to analyze spatial data which frequently occur in practice such as in health sciences and life studies. It is customary to incorporate spatial random effects into the model to account for spatial variation of the data. In particular, Poisson mixed models are used to analyze the spatial count data. It is often assumed that the observations in each area, conditional on the spatial random effects, are independent to each other. However, this may not be a valid assumption in practice. For instance, multiple asthma visits by a child to physicians (within a year) are not clearly independent observations. To address this issue, this paper develops spatial models with repeated events. In particular, compound Poisson mixed models are introduced to account for the repeated events as well as the spatial variation of the data. Performance of the proposed approach is evaluated through simulation studies and by a real dataset of children asthma visits to physicians in the province of Manitoba, Canada.

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

1. Introduction

The analysis of disease incidence (or mortality) over space has received considerable attention due to growing demand for a reliable disease mapping. The

*Mahmoud Torabi
Email address: `Mahmoud.Torabi@umanitoba.ca` (Mahmoud Torabi)

idea behind developments on spatial modelling of disease incidence is to essentially model variations in true disease pattern and better separate systematic variability from random noise, a component that usually overshadows crude disease incidence map. Map of areal disease incidence is a useful tool in determining spatial pattern of disease incidence for targeting resources. Disease incidence rates may differ substantially across geographical areas. A reliable estimate of the underlying disease risk is usually provided by borrowing strength from neighbouring geographic areas.

Poisson regression is commonly used for the analysis of disease cases, which implicitly assumes that the cases in nearby areas are independent and the variance of response is equal to the mean. However, these may not be reasonable assumptions because causal factors of the disease that are unmeasured or unknown and thus omitted from the regression model can lead 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 [1] 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 Intrinsic CAR model [2](Besag et al., 1991), Proper CAR model [3] (Besag, 1974) , and Leroux CAR (LCAR) model [4](Leroux et al., 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 [5, 6, 7, 8, 9]. The method of penalized quasi-likelihood (PQL) may also be used for inference in the GLMMs. This method was first proposed by Breslow and Clayton [10]

and they provided an example of the use of PQL for estimation in mapping studies. Monte Carlo Expectation-Maximization (MCEM) approach [11] and Monte Carlo Newton-Raphson (MCNR) algorithm [12] were also used for inference in the GLMMs. Lele *et al.* [13] introduced a frequentist approach, called data cloning (DC), to compute the maximum likelihood estimates (MLE) and their standard errors for general hierarchical models. Lele *et al.*[14] described an approach to compute prediction and prediction interval of random effects in the context of GLMMs. Recently, Torabi[15, 16] considered the use of DC method in the areas of spatial and spatio-temporal Poisson models, respectively.

As another alternative method, one may want to use the generalized estimating equation (GEE) approach proposed by Liang and Zeger [17] and Prentice and Zhao [18], to analyze longitudinal data using the generalized linear models (GLMs). Albert and Mcshane [19] and Lin and Clayton [20] studied the idea of GEE in the context of spatial GLMMs. However, the GEE was originally developed for longitudinal data, which assumes that the subjects (clusters) are independent from each other, so it can be applied to spatial data by neglecting some spatial dependency of data (e.g., Carl and Kuhn[21]). To overcome this issue, Lin and Clayton [20] considered the quasi likelihood (QL) approach to account for the full spatial covariance structure of the data. Torabi and Rosychuk [22] studied spatio-temporal 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 from 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 of repeated events, Tascheri *et al.* [23] studied the GLMM framework with sev-

eral candidate models including delta-lognormal, delta-gamma, quasi-Poisson and compound Poisson. Rosychuk *et al.* [24] introduced a compound Poisson method to account for repeated events under the frame of disease cluster detection. Taking into account the spatial models, Gschlobl and Czado [25] used a spatial Poisson model for the number of claims, while claim size was modeled using a gamma distribution. They incorporated the both covariates and spatial random effects into the model and used the Bayesian approach for the inference.

To the best of our knowledge, spatial modeling of repeated events has not been explored. The importance of considering repeated events in particular for chronic diseases with also accounting for spatial dependency in the context of disease mapping may led us to have a better understanding of spatial trend of disease and potential risk factors for possible preventions. To that end, we introduce a spatial discrete compound Poisson model to account for repeated events as well as 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 smoothed disease ratio in the context of spatial compound Poisson model in Section 4. In Section 5, performance of the proposed approach is evaluated using a real dataset of children 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 some concluding remarks are given in Section 7. Technical details are deferred to the Appendices.

2. Spatial compound Poisson model

Suppose that our population study is divided to m non-overlapping areas such as counties, provinces or municipalities, and outcome data are available as counts (e.g., number of disease cases in each area). Let C_{iw} be the random variable representing the number of individuals with exactly w events at area

$i(= 1, ..., m)$, and $w(< \infty)$ is the possible number of repetition (e.g., asthma visits to physicians). Then $C_i = \sum_w C_{iw}$ is the number of cases at area i. Let F_{ij} be the number of events of the jth individual ($j = 1, ..., C_i$) at area i. Hence, $Y_i = \sum_w wC_{iw} = \sum_{j=1}^{C_i} F_{ij}$ is the random variable representing total number of events at area i.

2.1. Statistical model and assumptions

Let F_{ij} follows 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

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

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 $\boldsymbol{\eta} = (\eta_1, ..., \eta_m)^\top$ represent spatial random effects. In particular, the LCAR model [4] is used to capture the spatial random effects $\boldsymbol{\eta}$. We consider the following general model for the spatial random effects $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta),$$

$$\boldsymbol{\Sigma}_\eta = \sigma_\eta^2 \big[(1 - \lambda_\eta)\mathbf{I_m} + \lambda_\eta \mathbf{R}\big]^{-1},$$

where $\mathbf{I_m}$ is the identity matrix of dimension m; $\mathbf{R}$ is a $m \times m$ intrinsic autoregressive matrix with elements $R_{ii} = w_i$ where w_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 neighbours; 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:

$$g[E(\boldsymbol{Y}|\boldsymbol{\eta})] = \text{offset} + \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{Y} = (Y_1, ..., Y_m)^\top$; $g(\cdot) = [\log(\cdot) - \log\lambda_w]$, (see Appendix A); the offset is the known vector of the logarithm of the e_i; the covariate matrix $\mathbf{x} = [\boldsymbol{J}, \{\{x_{ij}\}_{i=1}^m\}_{j=1}^p]$ 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 $m \times m$.

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 $\boldsymbol{\theta}_2 = (\lambda_\eta, \sigma_\eta^2)^\top$ 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 $Y_i, (i = 1, ..., m)$, we can write $\mu_i(\boldsymbol{\theta}) \equiv E(Y_i) = E[E(Y_i|\boldsymbol{\eta})] = \lambda_w \exp(\log e_i + \mathbf{x}_i^\top\boldsymbol{\beta}) M_\eta(\mathbf{z}_i)$, where $M_\eta(\mathbf{z}_i) = \exp(\mathbf{z}_i^\top\boldsymbol{\Sigma}_\eta\mathbf{z}_i/2) = \exp(\frac{1}{2}\Sigma_\eta^{ii})$ and Σ_η^{ii} is the ith diagonal element of $\boldsymbol{\Sigma}_\eta$. To get the marginal variance of $Y_i, (i = 1, ..., m)$, we can write

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

and,

$$\begin{aligned}\sigma_{ij}(\boldsymbol{\theta}) \equiv cov(Y_i, Y_j) =& \lambda_w^2 \exp\Big[\log e_i + \log e_j + (\mathbf{x}_i + \mathbf{x}_j)^\top\beta\Big]\Big\{ \exp[\frac{1}{2}(\Sigma_\eta^{ii} + \Sigma_\eta^{jj})] \\ &\times [\exp(\Sigma_\eta^{ij}) - 1]\Big\},\end{aligned} \tag{3}$$

(see Appendix B for derivation of $\mathbf{V}_1(\boldsymbol{\theta}) = cov(Y) = \{\sigma_{ij}(\theta)\}_{i,j=1}^m$).

3.1. Quasi-likelihood equations for fixed effects

We define $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, ..., \mu_m)^\top$ 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:

$$\mathbf{D}_1^\top(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)\mathbf{V}_1^{-1}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)[\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2)] = \mathbf{0}, \tag{4}$$

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 C 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 kth iteration, $\hat{\boldsymbol{\theta}}_1^{(k+1)}$ is obtained at the $(k+1)$th iteration as:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1^{(k+1)} \approx \hat{\boldsymbol{\theta}}_1^{(k)} + &\left[\mathbf{D}_1^\top(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\mathbf{V}_1^{-1}(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\mathbf{D}_1(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\right]^{-1} \\ &\times \left\{\mathbf{D}_1^\top(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\mathbf{V}_1^{-1}(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\left[\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2)\right]\right\}.\end{aligned} \tag{5}$$

One can then get the variance of $\hat{\boldsymbol{\theta}}_1$ (see Appendix D) estimated by

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

3.2. Quasi-likelihood equations for variance components

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

$$\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)[\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)] = \mathbf{0}, \tag{6}$$

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 F 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 kth iteration, $\hat{\boldsymbol{\theta}}_2^{(k+1)}$ is obtained at the $(k+1)$th iteration as:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_2^{(k+1)} \approx \hat{\boldsymbol{\theta}}_2^{(k)} &+ \left[\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})\mathbf{D}_2(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})\right]^{-1} \\ &\times \left\{\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})[\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)})]\right\}.\end{aligned} \tag{7}$$

Variance of $\hat{\boldsymbol{\theta}}_2$ can be also estimated by (see Appendix G for detail):

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

3.3. Complete algorithm to estimate the model parameters

The complete algorithm to estimate the model parameters $\boldsymbol{\theta}$ based on the QL approach using the equations (5) and (7) is given below:

1. Choose initial values $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$. Set $k = 0$.
2. Calculate $\hat{\boldsymbol{\theta}}_1^{(k+1)}$ using (5) and then $\hat{\boldsymbol{\theta}}_2^{(k+1)}$ using (7).
3. If convergence is reached, set $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_1^{(k+1)}$ and $\hat{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}}_2^{(k+1)}$; otherwise set $k = k + 1$ and return to step 2.

The estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ is based on unbiased estimating equations. Thus, it is not surprising that under some mild regularity conditions [26], it can be shown that this estimator is consistent. This estimator is special M-estimator [27] which follows asymptotic normality under some more regularity conditions. Our estimator is then fully efficient as we use the true variance-covariance matrix for the fixed effect and variance component parameters through $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively. Note that using a *working* variance-covariance matrix may lead to an inefficient estimator [28].

4. Disease ratio

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

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

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

$$\hat{\eta_i} = \widehat{E(\eta_i|Y_i)} = E(\eta_i|Y_i)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

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

$$E(\eta_i|Y_i) = \frac{\int_{-\infty}^{\infty} \eta_i P_{Y_i|\eta_i}(Y_i|\eta_i) f_{\eta_i}(\eta_i) d\eta_i}{\int_{-\infty}^{\infty} P_{Y_i|\eta_i}(Y_i|\eta_i) f_{\eta_i}(\eta_i) d\eta_i},$$

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

where $r = 0, 1$ and $K_i = e_i e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$. Hence, by Laplace approximation [29],we can estimate $E(\eta_i|Y_i)$ as:

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

where

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

and

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

where $\eta^*_{r,i}$ is the mode of $L_{r,i}(\eta_i)$ and can be obtained numerically through the *R software* [30] 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 the 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 asthma visits totalled 694,484 over the study period with mean and median number of yearly cases per area of 1,037 and 493 (range 36 to 5,718), respectively. The average population of children with the age under 18 years was about $302,949$. The areal child population sizes varied from 287 to 21,966, with mean and median numbers of 4,522 and 2,678, respectively. Figure 1 gives a histogram of number of visits for asthma to physicians by individuals (re-admissions) during the study period. 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 from each other is not a valid assumption.

We first fit the model (1) to the dataset of children asthma visits to physicians using the QL estimating equations (4) and (6). 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 ith 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. Table 1 reports the model parameter estimates and corresponding standard errors for the proposed model. Note that we estimate and report σ_η rather than σ_η^2 for simplicity in computation [4]. It seems that all the model parameters have significant contributions to the model. In particular, it appears that the estimation of λ_w is significantly larger than 1 which is the basic condition for 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

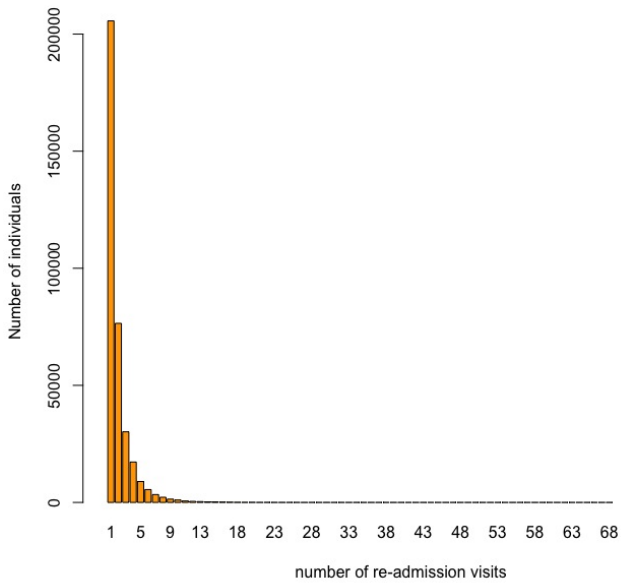

Figure 1: Histogram of number of childhood asthma visits by individuals in Manitoba, Canada, during 2000–2009.

each area, given the spatial random effects, is more explored in the simulation study (Section 6).

One of the main interests in spatial statistics is to also predict the disease ratio for each area and provide the corresponding map. Figure 2 presents map of the smoothed disease ratio (SDR) for children asthma visits to physicians in the province of Manitoba, Canada, during 2000—2009. The areas with SDR larger than one show higher number of childhood asthma visits compared to the rest of population. From Figure 2, it seems that south-east of the province have higher childhood asthma visits compared to the rest of population. These findings may represent real increases or different distributions of important covariates that are unmeasured and unadjusted for in our modeling.

We also provide map of prediction of spatial random effects in Figure 3. The spatial component resembles the map of disease ratio (Figure 2) which clearly

Table 1: Model parameter estimates (EST) and corresponding standard errors (SE) using spatial compound Poisson model for childhood asthma visits in Manitoba, Canada, during 2000 – 2009.

	Parameter	EST	SE
Fixed effects	β_0	-2.643	0.402
	λ_w	5.565	0.072
Variance components	λ_η	0.478	0.0007
	σ_η	1.959	0.0008

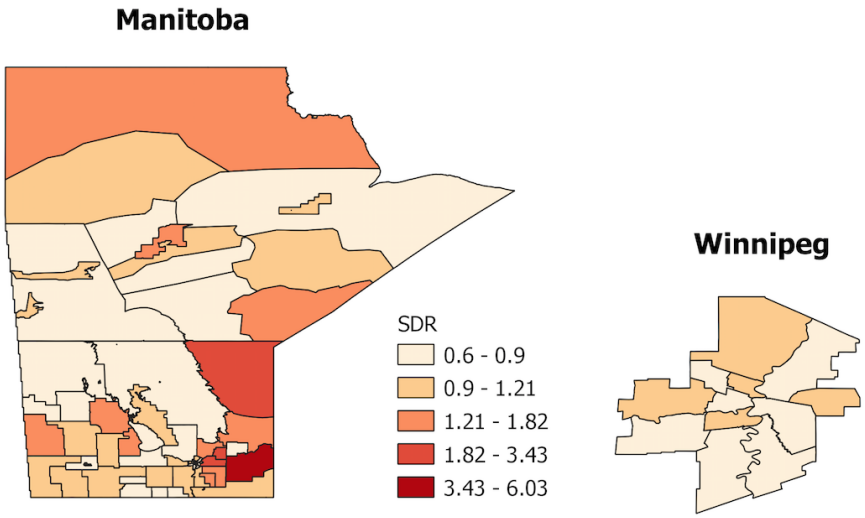

Figure 2: Prediction of disease ratio of childhood asthma visits in Manitoba, Canada, in 67 regional health authority districts.

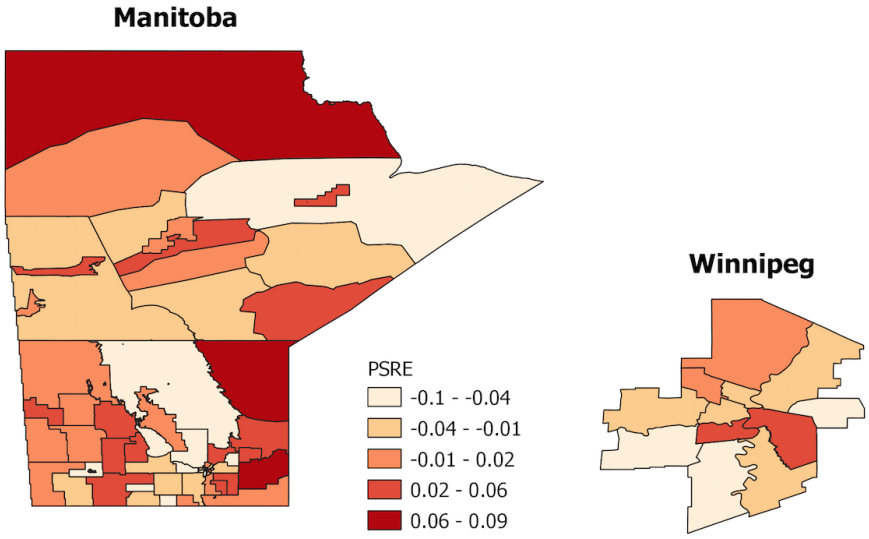

Figure 3: Prediction of spatial random effects (PSRE) of our childhood asthma visit dataset in Manitoba, Canada.

shows the dominance of a strong underlying spatial structure in our data.

6. Simulation study

We also conduct simulation studies to evaluate and compare performance of our proposed approach with the naive approach which ignores the dependence of observations in each area given the spatial random effects. 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 the both proposed and naive models. The Newton Raphson parameters are updated with maximum 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 ($\sigma_\eta \geq 0$, $0 \leq \lambda_\eta \leq 1$).

6.1. Irregular Grid

Here we use spatial layout of our Manitoba RHADs (Section 5), which is an irregular neighbourhood structure. Data are generated from the model (1) 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.

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) in the case of irregular grid.

		Proposed Model		Naive Model	
Parameter	True Value	Bias	MSE	Bias	MSE
β_0	-2.64	-0.461	1.039	1.791	3.214
λ_w	5.57	-0.076	0.034	—	—
λ_η	0.5	- 0.046	0.002	0.258	0.067
σ_η	2	-0.120	0.014	-0.481	0.231

We also provide boxplots (Figure 4) of model parameter estimates for the both proposed and naive models to have a better understanding of 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 from Figure 4 that the variability of estimated parameters in the case of proposed model is smaller than the naive model and in particular 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 compare 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 proposed model are considered.

With the same setting as used in Table 2, we set the parameter values

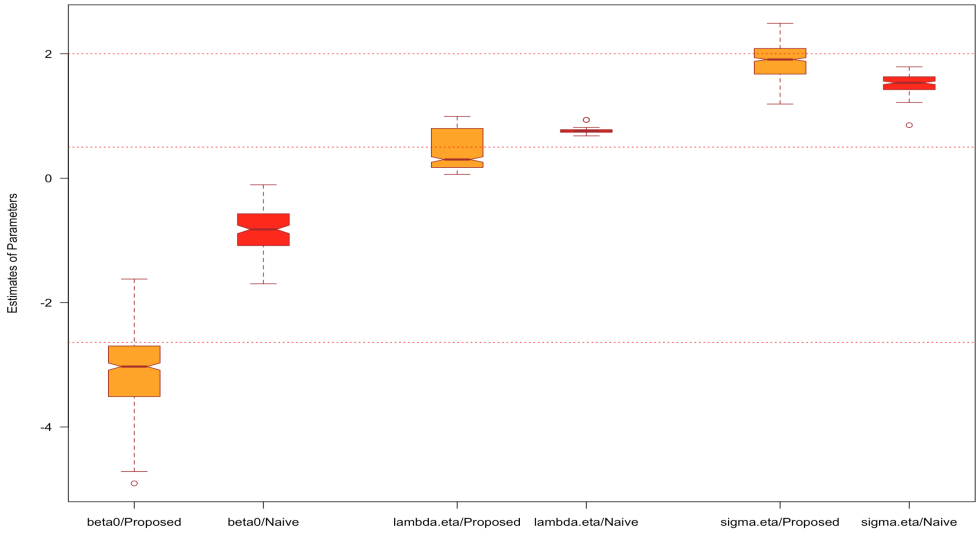

Figure 4: Boxplots of the model parameter estimates for the proposed and naive models in the case of irregular grid; dashed lines are true value lines.

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 proposed model are consistently smaller than the corresponding values in the case of naive model.

6.2. Regular Grid

In this part of simulation, we generate 10 by 10 lattice and calculate adjacency matrix through the *R software* [30] built-in function *Adjecancy.matrix*. Moreover, e_i's as offsets needed in model (1) are also generated uniformly between $[300, 320]$ and fixed during the simulation study. We set model parameters

Table 3: 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.

Model	True Parameter	β_0		λ_η		σ_η		λ_w	
		-2.5		0.5		2			
	λ_w	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Proposed	1	-0.207	0.147	-0.049	0.002	-0.060	0.004	-0.134	0.172
Naive		-0.569	1.300	-0.067	0.004	-0.129	0.017	—	—
Proposed	5	-0.473	0.906	-0.061	0.003	-0.124	0.015	-0.081	0.035
Naive		1.717	2.955	0.250	0.062	-0.503	0.253	—	—
Proposed	10	-0.509	1.002	-0.042	0.002	-0.141	0.020	-0.048	0.010
Naive		2.002	4.011	0.264	0.070	-0.584	0.341	—	—
Proposed	15	-0.468	1.209	-0.026	0.001	-0.171	0.029	-0.030	0.005
Naive		2.004	4.015	0.254	0.064	-0.574	0.329	—	—

as $\beta_0 = -5.5, \lambda_\eta = 0.8, \sigma_\eta = 2.5$ and four different values for $\lambda_w = 1, 5, 10$, and 15. We generate 500 datasets from the both proposed and naive models, and the model parameter estimates and MSEs are presented in Table 4. We also provide boxplots of model parameter estimates for the both proposed and naive models in the case of $\lambda_w = 15$ to also see the variation of estimated parameters (Figure 5).

As it is seen in Table 4, the biases and MSEs of the estimated parameters in the case of proposed model are smaller than the corresponding values of the naive model in most cases. Note that in the 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 5, the variability of estimated parameters in the case of proposed model is noticeably smaller than the naive model and in particular for spatial random effect parameters.

Table 4: 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 regular grid.

		β_0		λ_η		σ_η		λ_w	
Model	True Parameter	-5.5		0.8		2.5			
	λ_w	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Proposed	1	-0.213	0.201	-0.121	0.015	0.088	0.008	-0.266	0.354
Naive		-0.275	0.801	-0.010	0.0001	-0.003	0.00003	—	—
Proposed	5	-0.296	0.769	-0.015	0.0003	-0.000002	0.00004	-0.060	0.032
Naive		0.626	0.424	-0.166	0.028	-0.011	0.0001	—	—
Proposed	10	-0.401	0.906	-0.031	0.001	-0.016	0.0003	-0.040	0.009
Naive		0.980	0.968	-0.289	0.083	-0.003	0.0002	—	—
Proposed	15	-0.401	0.885	-0.027	0.001	-0.001	0.002	-0.027	0.004
Naive		0.995	0.994	-0.410	0.168	0.060	0.004	—	—

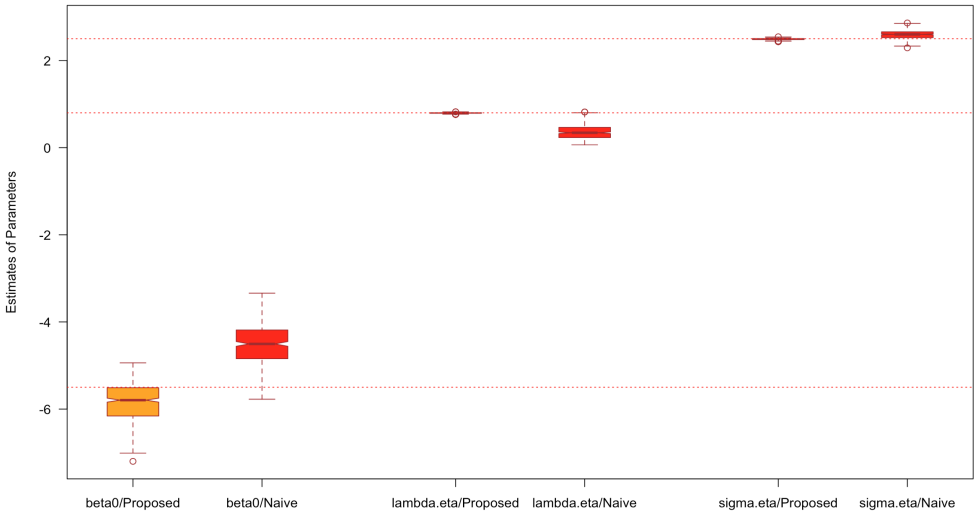

Figure 5: Boxplots of estimated parameters for the proposed and naive models when $\lambda_w = 15$ in the case of regular grid; dashed lines are true value lines.

7. Concluding remarks

In many applications, there are count data, for example, in an area, conditional on spatial random effects, that are not independent from each other. For instance, in our childhood asthma visit data application, we had in average almost 5 number of visits (range 1 to 68) as the 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 smoothed disease ratio over space was also provided.

We adjusted our expected number of children 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 spatio-temporal count data with repeated events. For instance, in our application, we may have access to children asthma visits to physicians data over space and each time visit for each child. Another interesting contribution would be to extend our proposed approach for binary data which has many applications [31]. These are some of the topics for future study.

Supplementary materials

The supplementary materials contain R codes and corresponding "readme" files for the simulations and real data application conducted in this paper.

Acknowledgements

We would like to thank the Editor and a referee for constructive comments and suggestions, which led to an improved version of the manuscript. This research is a part of PhD thesis work of Shabnam Balamchi. Mahmoud Torabi gratefully acknowledges the research support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Shabnam Balamchi's research was supported by Graduate Enhancement of Tri-Council Stipends (GETS) at the University of Manitoba. We also thank Manitoba Centre for Health Policy (MCHP) for providing the Asthma data set.

Appendix

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

A Derivation of the link function as $\{\log(\cdot) - \log \lambda_w\}$

We can write

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

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

$$E(Y_i|\boldsymbol{\eta}) = \lambda_w \exp(\text{offset} + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}).$$

We then have

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

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

B Derivation of $\mathbf{V}_1(\boldsymbol{\theta})$

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

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

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

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

C *Derivation of* $\mathbf{D}_1(\boldsymbol{\theta})$

To obtain $\mathbf{D}_1(\boldsymbol{\theta}) = \big\{\mathbf{D}_{1i}(\boldsymbol{\theta})\big\}_{i=1}^m$, 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 $\partial\mu_i(\boldsymbol{\theta})/\partial\lambda_w$, $(i = 1, ..., m)$. To that end, we can write

$$
\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} = \mathbf{x}_\mathbf{i}^\top[\lambda_\mathbf{w} \exp(\log \mathbf{e_i} + \mathbf{x}_\mathbf{i}^\top \boldsymbol{\beta})\mathbf{M}_\eta(\mathbf{z_i})] = \mathbf{x}_\mathbf{i}^\top \mu_\mathbf{i},
$$

and

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

315

D *Derivation of* $var(\hat{\boldsymbol{\theta}}_1)$

By applying Taylor expansion and using equation (4), we can write

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

Moreover, since

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

we then have

$$var(\hat{\boldsymbol{\theta}}_1)\approx\Big[\mathbf{D}_1^{\top}(\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2)\mathbf{V}_1^{-1}(\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2)\mathbf{D}_1(\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2)\Big]^{-1}.$$

E Derivation of $\mathbf{V}_2(\boldsymbol{\theta})$

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

$$\mathcal{V}_{ii}=var(S_i)=E(Y_i-\mu_i)^4-E^2(Y_i-\mu_i)^2$$
$$=E(Y_i-\mu_i)^4-{\sigma_{ii}}^2.$$

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

$$E(Y_i-\mu_i)^4=E(Y_i^4)-4\mu_iE(Y_i^3)+6\mu_i^2E(Y_i^2)-3\mu_i^4,$$

where

$$E(Y_i^4)=E[E(Y_i^4|\boldsymbol{\eta})]=(1+7\lambda_w+6\lambda_w^2+\lambda_w^3)E(\mu_{i\eta})+(7+18\lambda_w+7\lambda_w^2)E({\mu_{i\eta}}^2)$$
$$+6(1+\lambda_w)E({\mu_{i\eta}}^3)+E({\mu_{i\eta}}^4),$$

$$E(Y_i^3)=E[E(Y_i^3|\boldsymbol{\eta})]=(1+3\lambda_w+\lambda_w^2)E(\mu_{i\eta})+3(1+\lambda_w)E({\mu_{i\eta}}^2)+E({\mu_{i\eta}}^3),$$

and

$$E(Y_i^2)=E[E(Y_i^2|\boldsymbol{\eta})]=(1+\lambda_w)E(\mu_{i\eta})+E({\mu_{i\eta}}^2).$$

Hence,

$$E(Y_i-\mu_i)^4=\mu_i(1+7\lambda_w+6\lambda_w^2+\lambda_w^3)-4\mu_i^2(1+3\lambda_w+\lambda_w^2)+6\mu_i^3(1+\lambda_w)-3\mu_i^4$$
$$+\Big[(7+18\lambda_w+7\lambda_w^2)-12\mu_i(1+\lambda_w)+6\mu_i^2\Big]E({\mu_{i\eta}}^2)$$
$$+2\Big[3(1+\lambda_w)-2\mu_i\Big]E({\mu_{i\eta}}^3)+E({\mu_{i\eta}}^4),$$

where

$$E(\mu_{i\eta}) = \mu_i = \lambda_w \exp(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}) M_\eta(\mathbf{z}_i),$$

$$E({\mu_{i\eta}}^2) = \lambda_w^2 \exp\Big[2(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta})\Big] M_\eta(2\mathbf{z}_i),$$

$$E({\mu_{i\eta}}^3) = \lambda_w^3 \exp\Big[3(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta})\Big] M_\eta(3\mathbf{z}_i),$$

$$E({\mu_{i\eta}}^4) = \lambda_w^4 \exp\Big[4(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta})\Big] M_\eta(4\mathbf{z}_i).$$

Finally, we can write

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

where

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

F Derivation of $\mathbf{D}_2(\boldsymbol{\theta})$

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

$1, ..., m$). To that end, by using the equations (2) and (3), we can write

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

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

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

G *Derivation of* $var(\hat{\boldsymbol{\theta}}_2)$

By applying Taylor expansion and using equation (6), we have

$$\hat{\boldsymbol{\theta}}_2 \approx \boldsymbol{\theta}_2 + \Big[\mathbf{D}_1^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{D}_2(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\Big]^{-1}$$
$$\times \Big\{\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)[\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)]\Big\}.$$

Moreover, since

$$var\Big\{\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)[\boldsymbol{S}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)]\Big\}$$
$$= \Big[\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{D}_2(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\Big],$$

we can then write

$$var(\hat{\boldsymbol{\theta}}_2) \approx \Big[\mathbf{D}_2^\top(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{V}_2^{-1}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\mathbf{D}_2(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2)\Big]^{-1}.$$

References

[1] D. Clayton, J. Kaldor, Empirical bayes estimates of age-standardized relative risks for use in disease mapping, Biometrics (1987) 671–681.

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

[3] J. Besag, Spatial interaction and the statistical analysis of lattice systems, Journal of the Royal Statistical Society: Series B (Methodological) 36 (2) (1974) 192–225.

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

[5] L. Bernardinelli, C. Montomoli, Empirical bayes versus fully bayesian analysis of geographical variation in disease risk, Statistics in medicine 11 (8) (1992) 983–1007.

[6] D. G. Clayton, Bayesian methods for mapping disease risk, Geographical and environmental epidemiology: methods for small-area studies (1992) 205–220.

[7] L. A. Waller, B. P. Carlin, H. Xia, A. E. Gelfand, Hierarchical spatio-temporal mapping of disease rates, Journal of the American Statistical association 92 (438) (1997) 607–617.

[8] L. Knorr-Held, Bayesian modelling of inseparable space-time variation in disease risk, Statistics in medicine 19 (17-18) (2000) 2555–2567.

[9] M. Torabi, R. J. Rosychuk, Hierarchical bayesian spatiotemporal analysis of childhood cancer trends, Geographical Analysis 44 (2) (2012) 109–120.

[10] N. E. Breslow, D. G. Clayton, Approximate inference in generalized linear mixed models, Journal of the American statistical Association 88 (421) (1993) 9–25.

[11] C. E. McCulloch, Maximum likelihood variance components estimation for binary data, Journal of the American Statistical Association 89 (425) (1994) 330–335.

[12] C. E. McCulloch, Maximum likelihood algorithms for generalized linear mixed models, Journal of the American statistical Association 92 (437) (1997) 162–170.

[13] S. R. Lele, B. Dennis, F. Lutscher, Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods, Ecology letters 10 (7) (2007) 551–563.

[14] S. R. Lele, K. Nadeem, B. Schmuland, Estimability and likelihood inference for generalized linear mixed models using data cloning, Journal of the American Statistical Association 105 (492) (2010) 1617–1625.

[15] M. Torabi, Spatial modeling using frequentist approach for disease mapping, Journal of Applied Statistics 39 (11) (2012) 2431–2439.

[16] M. Torabi, Spatio–temporal modeling for disease mapping using car and b-spline smoothing, Environmetrics 24 (3) (2013) 180–188.

[17] K.-Y. Liang, S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1) (1986) 13–22.

[18] R. L. Prentice, L. P. Zhao, Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, Biometrics (1991) 825–839.

[19] P. S. Albert, L. M. McShane, A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data, Biometrics (1995) 627–638.

[20] P.-S. Lin, M. K. Clayton, et al., Analysis of binary spatial data by quasi-likelihood estimating equations, The Annals of Statistics 33 (2) (2005) 542–555.

[21] G. Carl, I. Kühn, Analyzing spatial autocorrelation in species distributions using gaussian and logit models, ecological modelling 207 (2-4) (2007) 159–170.

[22] M. Torabi, R. J. Rosychuk, Spatio-temporal modelling of disease mapping of rates, Canadian Journal of Statistics 38 (4) (2010) 698–715.

[23] R. Tascheri, J. Saavedra-Nievas, R. Roa-Ureta, Statistical models to standardize catch rates in the multi-species trawl fishery for patagonian grenadier (macruronus magellanicus) off southern chile, Fisheries Research 105 (3) (2010) 200–214.

[24] R. J. Rosychuk, C. Huston, N. G. Prasad, Spatial event cluster detection using a compound poisson distribution, Biometrics 62 (2) (2006) 465–470.

[25] S. Gschlößl, C. Czado, Spatial modelling of claim frequency and claim size in non-life insurance, Scandinavian Actuarial Journal 2007 (3) (2007) 202–225.

[26] C. C. Heyde, Quasi-likelihood and its application: a general approach to optimal parameter estimation, Springer Science & Business Media, 2008.

[27] J. Wooldridge, M.(2002) econometric analysis of cross section and panel data, The MIT Press 0 5 (1) (1995) 5.

[28] G. M. Fitzmaurice, A caveat concerning independence estimating equations with multivariate binary data, Biometrics (1995) 309–317.

[29] L. Tierney, J. B. Kadane, Accurate approximations for posterior moments and marginal densities, Journal of the american statistical association 81 (393) (1986) 82–86.

[30] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2013).
URL http://www.R-project.org/

[31] M. Torabi, Spatiotemporal modeling of odds of disease, Environmetrics 25 (5) (2014) 341–350.