Spatio-temporal modeling of disease mapping of rates

Key words and phrases: Conditional autoregressive; generalized estimating equation; generalized linear mixed model; geographic epidemiology; random effects model; seasonal effect.

Abstract: This paper studies generalized linear mixed models (GLMMs) for the analysis of geographic and temporal variability of disease rates. This class of models adopts spatially correlated random effects and random temporal components. Spatio-temporal models that use conditional autoregressive smoothing across the spatial dimension and autoregressive smoothing over the temporal dimension are developed. The model also accommodates the interaction between space and time. However, the effect of seasonal factors has not been previously addressed and in some applications (e.g. health conditions), these effects may not be negligible. We incorporate the seasonal effects of month and possibly year as part of the proposed model and estimate model parameters through generalized estimating equations. The model provides smoothed maps of disease risk and eliminates the instability of estimates in low-population areas while maintaining geographic resolution. We illustrate our approach using a monthly data set of the number of asthma presentations made by children to Emergency Departments (EDs) in the province of Alberta, Canada during the period 2001-2004.

1. INTRODUCTION

The analysis of disease rates over space and time has received considerable attention due to growing demand for reliable disease rates. The idea behind developments on spatial and spatio-temporal modeling of disease rates is essentially to model variations in true rates and better separate systematic variability from random noise, a component that usually overshadows crude rate maps. Maps of regional morbidity and mortality rates over time are useful tools in determining spatial and temporal patterns of disease. Disease incidence and mortality rates may differ substantially across geographical regions. A reliable estimate of the underlying disease risk is usually provided by *borrowing strength* from neighbouring geographic sub-areas.

Poisson regression is commonly used for the analysis of disease rates, which implicitly assumes that the rates in nearby regions 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 regions. Clayton & Kaldor (1987) extended the use of mixed models for geographical data to account for the extra-Poisson variability through the introduction of random effects; where the random effects are often spatially correlated in a disease mapping context. To perform inference based on mixed models, computationally intensive hierarchical models are commonly used. One may use Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Besag, York & Molliè 1991; Clayton & Bernardineli 1992; Waller, Carlin, Xia & Gelfand 1997; Knorr-Held 2000), but monitoring the algorithm for convergence is difficult (Bernardineli & Montomoli 1992; Breslow & Clayton 1993). Breslow & Clayton (1993) proposed the use of the penalized quasi-likelihood (PQL) method for inference in generalized linear mixed models (GLMMs) and provided an example of the use of PQL for estimation in mapping studies. For the analysis of spatial rates, PQL may require more computer time than the Gibbs sampler, due to the need for large-scale matrix computations. Liang & Zeger (1986) and Prentice & Zhao (1991) proposed the generalized estimating equation (GEE) approach to analyzing the longitudinal data using the generalized linear models.

Waller, Carlin, Xia & Gelfand (1997) extended the existing Bayesian hierarchical spatial models to account for temporal effects and spatio-temporal interactions. More precisely, they proposed spatio-temporal models for mapping rates, including temporal effects, through an autoregressive AR(1) structure. Knorr-Held (2000) proposed a unified approach for a Bayesian analysis of incidence or mortality data in space and time which has four different types of prior distributions for the interaction of space and time; this constitutes an extension of a model with only main effects. The random walk RW(1) with independent Gaussian increments was used as a temporal component in the model proposed by Knorr-Held (2000). MacNab & Dean (2001) proposed spatio-temporal models that use autoregressive local smoothing across the spatial dimension and B-spline smoothing over the temporal dimension. In some contexts, the underlying rates may change over seasons within a given year. For example, in the infectious disease context, malaria (Mabaso, Craig, Vounatsou & Smith 2005) and influenza-related mortality (Greene, Ionides & Wilson 2006) have been noted to have spatio-temporal as well as seasonal effects. Such seasonal effects are likely to be important. Existing temporal smoothing techniques are not applicable in this context and new model is required to handle seasonal effects.

We consider a comprehensive model that is based on a generalized linear mixed model (GLMM) to account for the spatio-temporal analysis of risks. These models accommodate spatially correlated random effects as well as temporal effects. The well-known approaches of conditional autoregressive (CAR) and AR models are adopted for spatial and temporal random effects, respectively. We use the space-time interaction to capture any additional effects that are not explained by the main factors of space and time. The seasonal effects resulting from the effects of month and possibly year are incorporated in our model.

The paper is organized as follows. In Section 2, the spatio-temporal GLMM is studied. We provide the statistical modeling of spatial and temporal factors as well as the interaction between space and time. Furthermore, we study the seasonal factors of month and year. The estimation of model parameters, GEE approach, is described in detail in Section 3. Section 4 illustrates the model using the number of asthma visits made by children to EDs in the province of Alberta, Canada during 2001 to 2004. The performance of the GEE approach in our application is also studied through the simulation study. Concluding remarks are given in Section 5. The Appendix provides some of the calculation details.

2. SPATIAL AND TEMPORAL MODELING

2.1. Statistical models.

Let y_{it} be the counts of disease (or otherwise) for the *i-th* geographic area at time t and let n_{it} be the corresponding population at risk, for $i = 1, ..., I$, and $t = 1, ..., T$. Define μ^c_{it} as the conditional expectation of y_{it} given the random effects. We study a general Poisson model for μ^c_{it} which is given by

$$\mu^c_{it} = \exp\left\{\log n_{it} + \log \mu + A\cos(\pi 6/t) + B\sin(\pi 6/t) + \eta_i + \alpha_t + \theta_{it}\right\}, \quad (1)$$

where μ is a fixed effect representing the overall mean count over time and area. The effects of seasons are studied by $A\cos(\pi t/6) + B\sin(\pi t/6)$. The expressions $\cos(\pi t/6)$ and $\sin(\pi t/6)$ account for the seasonal variation over time with the coefficients A and B, respectively (Fanshawe, Diggle, Rushton, Lurz, Glinianaia, Pearce, Parker, Chalton & Pless-Mulloli 2008). This formulation allows the seasonal pattern to change over time, in contrast with simpler models where the components must sum exactly to 0 (Datta, Lahiri, Maiti & Lu 1999). As an alternative way to introduce the seasonal variation, we could apply the seasonal components $(s_1, ..., s_{11})$ proposed by Knorr-Held & Richardson (2003) such that the components add up to Gaussian white noise over a moving window of length 12 months. The η_i accounts for the spatial random effects. Moreover, α_t represents unspecified features of time t which displays temporal structure. To account for the interaction between space and time, motivated by our application, we provide an interaction effect θ_{it} in (1).

2.2. Spatial modeling.

The usual CAR model is used to capture the spatial random effects η_i. A variety of CAR models may be used by taking a collection of mutually compatible conditional distributions $p(\eta_i|\eta_{-i}), i = 1, ..., I$ where $\eta_{-i} = \{\eta_j : j \neq i, j \in \partial_i\}$ and ∂_i denotes a set of neighbours for the *i-th* area (Besag, York & Molliè 1991). We consider the following general model for the spatial effects η_i,

$$\eta = (\eta_1, ..., \eta_I)' \sim N(0, \Sigma_\eta), \tag{2}$$

$$\Sigma_\eta = \sigma_\eta^2 P^{-1},$$

$$P = \lambda_\eta D + (1 - \lambda_\eta) I_I,$$

where σ_η^2 is the spatial dispersion parameter, λ_η measures the spatial autocorrelation, $0 \leq \lambda_\eta \leq 1$, and I_I is an identity matrix of dimension I. The neighbourhood matrix D has its *i-th* diagonal element equal to the number of neighbours of the corresponding area ($\#\partial_i$), and the off-diagonal elements in each row equal -1 if the corresponding areas are neighbours and zero otherwise (Leroux, Lei & Breslow 1999; MacNab & Dean 2000, 2001). Neighbours can be defined in various ways, depending on the context of the analysis, but one popular definition is simply the set of areas that have common borders and we use this definition. When the spatial correlation λ_η is one, we have

$\eta_i|\eta_{-i} \sim N(\bar{\eta}, \sigma^2_\eta/\#\partial_i)$, where $\bar{\eta}$ is the mean of the random effects in the neighbourhood of the *i-th* area. A more general form of the spatial random effects η_i is given by

$$\eta_i|\eta_{-i} \sim N(\frac{\sum_{j\in\partial_i} w_{ij}\eta_j}{\sum_{j\in\partial_i} w_{ij}}, \frac{\sigma^2_\eta}{\sum_{j\in\partial_i} w_{ij}}), \tag{3}$$

where w_{ij} is the user-specified weights linking areas i and j (Wakefield & Morris 1999; Lawson, Biggeri, Böhning, Lesaffre, Viel & Bertollini 1999; MacNab & Dean 2000, 2001).

2.3. Temporal modeling.

One may employ the RW(1) for the random effects α_t such that $\alpha_t|\alpha_{t-1} \sim N(\alpha_{t-1}, \sigma^2_\alpha)$ with $\alpha_1 = 0$, where σ^2_α is an unknown scalar to be estimated from the data (Knorr-Held 2000; Gössl, Auer & Fahrmeir 2001). Then $\alpha = (\alpha_1, ..., \alpha_T)' \sim N(0, \sigma^2_\alpha K_t^{-1})$, where K_t is a known structure matrix as:

$$K_t = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

Because the specification is in terms of conditional distributions, K_t is a singular matrix. In generalized linear modeling, the typical solution would be to reduce the problem to that of estimating a reduced set of random effects with full rank variance matrix; in this case one can simply delete the first row and column of K_t (Clayton 1996). This can be equivalently handled using the Moore-Penroze generalized inverse K_t^- (Harville 1997).

Because of simplicity in the analytical work to derive the marginal mean and variance (see Section 3) and to capture all temporally correlated random effects, we use the AR(1) model for the random effects α_t. The simple AR(1) model for the random effects α_t may be written as $\alpha_t|\alpha_{t-1} \sim N(\rho\alpha_{t-1}, \sigma_\alpha^2)$, where σ_α^2 is the temporal dispersion parameter and ρ as temporal autocorrelation with $|\rho| \leq 1$ (Waller et al. 1997; Martínez-Bereito, López-Quilez & Botella-Rocamoraet 2008). The vector $\alpha = (\alpha_1, ..., \alpha_T)'$ is assumed to be multivariate Gaussian with mean zero and covariance matrix $\frac{\sigma_\alpha^2}{1-\rho^2}\{\rho^{|j-i|}\}_{i,j=1}^T$. The seasonal effects $\cos(\pi t/6)$ and $\sin(\pi t/6)$ are also used to account for the seasonal variation from month to month.

2.4. Spatio-temporal interaction.

The interaction effect of space and time θ_{it} may be defined in many different ways. One way to define θ_{it} is

$$\theta = (\theta_{11}, ..., \theta_{IT})' \sim N(0, \sigma_\theta^2 K_\theta^{-1}),$$

where σ_θ^2 measures the dispersion between space and time effects and K_θ is a pre-specified structure matrix. Note that if all $\theta_{it} = 0$, the model (1) reduces to model with the main effects of space and time. Hence, θ captures only the variation that is not explained by the main effects. Clayton (1996) suggests specifying K_θ as the Kronecker product of the structure matrices of those main effects, which are assumed to interact. Alternatively, one may define θ_{it} as $\theta_i t$ or $S_i(t)$ depending on the nature of the data (MacNab & Dean 2001; Silva, Dean, Niyonsenga & Vanasse 2008), where θ_i is a fixed parameter or an CAR model, and $S_i(t)$ is a cubic B-spline for specific region i (Eilers & Marx 1996). In this paper, we use $\theta = (\theta_{11}, ..., \theta_{IT})' \sim N(0, \sigma_\theta^2 K_\theta^{-1})$ which was found useful in our exploration of the data.

2.5. Full model.

The model (1) can be written as a GLMM, $E(y|v) = g(\text{offset} + X\beta + Zv)$, where $y = (y_{11}, ..., y_{IT})'$; $g(\cdot) = \exp(\cdot)$; the offset is the known vector of the logarithm of the population counts n_{ij}; $\beta = (\log\mu, A, B)'$ is the vector of fixed parameters; X and Z are the known $N \times \mathrm{p}$ and $N \times \mathrm{h}$ matrices of full rank, $(N = I \times T, \mathrm{p} = 3, \mathrm{h} = T + I + N)$; and $v = (\alpha_1, ...\alpha_T, \eta_1, ..., \eta_I, \theta_{11}, ..., \theta_{IT})'$ is independently distributed with mean 0 and covariance matrix $\Sigma_v = diag(\Sigma_\alpha, \Sigma_\eta, \sigma_\theta^2 K_\theta^{-1})$ depending on variance parameters $\varsigma = (\lambda_\eta, \sigma_\eta^2, \sigma_\alpha^2, \rho, \sigma_\theta^2)'$. The design matrix vector $X = col_{1\leq i\leq I}(X_i)$ corresponds to the fixed effects and has dimension $N \times \mathrm{p}$, where X_i is the same for all areas with the t-*th* row of X_i as $X_{it} = (1, \cos(\pi t/6), \sin(\pi t/6)), t = 1, ..., T$. The design matrix $Z = col_{1\leq i\leq I}(Z_i)$ for the random effects has dimension $N \times \mathrm{h}$, where $Z_i = (Z_0, Z_{i1}, Z_{i2}), i = 1, ..., I$. The matrix Z_0 is an $T \times T$ identity matrix, and Z_{i1} has dimension $T \times I$ where the corresponding i-*th* column is one, elsewhere 0; Z_{i2} has dimension $T \times N$ where it decomposes to I matrix with each dimension $T \times T$ where i-*th* matrix is an $T \times T$ identity matrix, elsewhere 0. Note that Z_0 is associated with α_t; Z_{i1} and Z_{i2} are associated with η_i and θ_{it}, respectively.

There are various methods to estimate the fixed parameters β and variance components ς such as generalized estimating equations (Liang & Zeger 1986; Prentice & Zhao 1991), penalized quasi-likelihood (Breslow & Clayton 1993; Lin & Breslow 1996; Leroux, Lei & Breslow 1999; MacNab & Dean 2000, 2001), estimating functions (Yasui & Lele 1997), hierarchical likelihood (Lee & Nelder 1996), Bayesian analysis using Markov chain Monte Carlo (Besag, York & Molliè 1991; Bernardinelli & Montomoli 1992; Gilks, Richardson & Spiegelhalter 996; Clayton & Bernardinelli 1992), and the EM algorithm (McCulloch 1997). We use the method of GEE to estimate the model parameters. However, our set-up is different from the classical set-up that was first proposed by Liang & Zeger (1986) and Prentice & Zhao (1991) and is further described in the next section.

3. ESTIMATION OF PARAMETERS

3.1. Generalized estimating equations for fixed effects.

We first need to find the marginal mean and marginal variance-covariance of y. Recall that $E(y|v) = \exp(\text{offset} + X\beta + Zv)$. Then for $i = 1, ..., I; t = 1, ..., T$, we have

$$\mu_{it}(\beta, \varsigma) := E(y_{it}) = E\Big\{E(y_{it}|v)\Big\} = \exp(\log n_{it} + x_{it}^{'}\beta)M_v(z_{it}), \qquad (4)$$

where $M_v(z_{it}) = \exp\left(z_{it}^{\prime}\Sigma_v z_{it}/2\right) = \exp\left\{\frac{1}{2}(\sigma_\theta^2 q_{it}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\right\}$; x_{it}

and z_{it} are the t-th row of X_i and Z_i, respectively; q_{it}^{-1} is the (i,t)-th element of K_θ^{-1}; p_{ii}^{-1} is the i-th diagonal element of P^{-1}.

To obtain the marginal variance of y_{it}, we may write

$$\begin{aligned}
\mathrm{var}(y_{it}) &= \mathrm{var}\Big\{E(y_{it}|v)\Big\} + E\Big\{\mathrm{var}(y_{it}|v)\Big\} \\
&= \mathrm{var}\Big\{\exp(\log n_{it} + x_{it}^{'}\beta + z_{it}^{'}v)\Big\} + E\Big\{\exp(\log n_{it} + x_{it}^{'}\beta + z_{it}^{'}v)\Big\} \\
&= \exp\Big\{2(\log n_{it} + x_{it}^{'}\beta)\Big\}\Big[M_v(2z_{it}) - \{M_v(z_{it})\}^2 + \exp(-\log n_{it} - x_{it}^{'}\beta)M_v(z_{it})\Big],
\end{aligned}$$

where $M_v(2z_{it}) = \exp\left\{2(\sigma_\theta^2 q_{it}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\right\}$. Therefore, we have

$$\begin{aligned}
\mathrm{var}(y_{it}) &= \mu_{it}\Bigg(\exp(\log n_{it} + x_{it}^{'}\beta)\Big[\exp\Big\{\frac{3}{2}(\sigma_\theta^2 q_{it}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\Big\} \\
&\quad - \exp\Big\{\frac{1}{2}(\sigma_\theta^2 q_{it}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\Big\}\Big] + 1\Bigg) \\
&= :\sigma_{itt}(\beta,\varsigma); i = 1,...,I; t = 1,...,T. \qquad (5)
\end{aligned}$$

Similarly, we may write $\mathrm{cov}(y_{it}, y_{it'})$ as

$$\begin{aligned}
\mathrm{cov}(y_{it}, y_{it'}) &= \exp\Big\{\log n_{it} + \log n_{it'} + (x_{it} + x_{it'})^{'}\beta\Big\}\Big[\exp\Big\{\sigma_\theta^2(q_{it}^{-1} + q_{it'}^{-1}) + \sigma_\alpha^2\frac{1+\rho^{|t-t'|}}{1-\rho^2} \\
&\quad + 2\sigma_\eta^2 p_{ii}^{-1}\Big\} - \exp\Big\{\sigma_\theta^2(q_{it}^{-1} + q_{it'}^{-1})/2 + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1}\Big\}\Big] \\
&= :\sigma_{itt'}(\beta,\varsigma); t \neq t' = 1,...,T. \qquad (6)
\end{aligned}$$

Moreover, we need to find $\mathrm{cov}(y_{it}, y_{jt}) =: \sigma_{ijtt}(\beta,\varsigma)$ and $\mathrm{cov}(y_{it}, y_{jt'}) =: \sigma_{ijtt'}(\beta,\varsigma)$ to construct the covariance between y_i and y_j ($i \neq j = 1,...,I; t \neq t' = 1,...,T$). We may write $\mathrm{cov}(y_{it}, y_{jt})$ and $\mathrm{cov}(y_{it}, y_{jt'})$ after some simplification as

$$\sigma_{ijtt}(\beta,\varsigma) = \text{cov}(y_{it},y_{jt}) = \exp\Big\{\log n_{it} + \log n_{jt} + (x_{it}+x_{jt})'\beta\Big\}\exp\Big\{\sigma_\theta^2(q_{it}^{-1}+q_{jt}^{-1})/2 + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2(p_{ii}^{-1}+p_{jj}^{-1})/2\Big\}\Big[\exp\{\sigma_\theta^2(q_{it}^{-1}+q_{jt}^{-1})/2 + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ij}^{-1}\} - 1\Big], \quad (7)$$

and

$$\sigma_{ijtt'}(\beta,\varsigma) = \text{cov}(y_{it},y_{jt'}) = \exp\Big\{\log n_{it} + \log n_{jt'} + (x_{it}+x_{jt'})'\beta\Big\}\exp\Big\{\sigma_\theta^2(q_{it}^{-1}+q_{jt'}^{-1})/2 + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2(p_{ii}^{-1}+p_{jj}^{-1})/2\Big\}\Big[\exp\{\sigma_\theta^2(q_{it}^{-1}+q_{jt'}^{-1})/2 + \sigma_\alpha^2\rho^{|t-t'|}/(1-\rho^2) + \sigma_\eta^2 p_{ij}^{-1}\} - 1\Big]. \quad (8)$$

We define $V_{ii}^{(1)}(\beta,\varsigma) = \text{cov}(y_i) = \{\sigma_{ikl}(\beta,\varsigma)\}_{k,l=1}^T$ and $V_{ij}^{(1)}(\beta,\varsigma) = \text{cov}(y_i,y_j) = \{\sigma_{ijkl}(\beta,\varsigma)\}_{k,l=1}^T, (i \neq j = 1,...,I)$. Hence, $V_1(\beta,\varsigma) = \text{cov}(y) = \{V_{ij}^{(1)}\}_{i,j=1}^I$. Then $\mu(\beta,\varsigma) = (\mu_{11},...,\mu_{IT})'$ is the mean vector of the response vector $y = (y_{11},...,y_{IT})'$ and $V_1(\beta,\varsigma)$ is the $N \times N$ variance-covariance matrix of y. Note that μ and V_1 are functions of β and ς. In a classical set-up, where μ and V_1 are functions of β only, one estimates β by solving the well-known quasi-likelihood estimating equations

$$\frac{\partial\mu'}{\partial\beta}V_1^{-1}(y-\mu) = 0,$$

(Wedderburn 1974; McCullagh 1983). Moreover, Liang & Zeger (1986) introduced the GEE approach where V_1 is a function of β as well as ς. However, in the present set-up, μ and V_1 involve unknown parameters β and ς. In the longitudinal fixed model set-up, Sutradhar & Das (1999) (also Jowaheer & Sutradhar 2002) proposed a generalized estimating approach that uses the consistent estimates of ς involved in the μ and V_1 matrix (see also Bari & Sutradhar 2005). In our spatio-temporal model, we use the $I^{1/2}-$consistent estimates of ς, $\hat{\varsigma}(y, \beta)$, involved in the μ and V_1, and solve the estimating equation with respect to fixed parameters β by using (4)-(8) which is given by

$$D_1'(\beta, \hat{\varsigma})V_1^{-1}(\beta, \hat{\varsigma})\{y - \mu(\beta, \hat{\varsigma})\} = 0, \tag{9}$$

where $D_1(\beta, \varsigma) = \partial\mu(\beta, \varsigma)/\partial\beta = (x_{11}\mu_{11}, ..., x_{1T}\mu_{1T}, ..., x_{I1}\mu_{I1}, ..., x_{IT}\mu_{IT})'$. The solution of equation (9), $\hat{\beta}$, may be obtained using the Newton-Raphson iterative method. Given the value $\hat{\beta}_s$ at the *s-th* iteration, $\hat{\beta}_{s+1}$ may be obtained at the *(s + 1)-th* iteration as

$$\begin{aligned} \hat{\beta}_{s+1} &= \hat{\beta}_s + \left\{D_1'(\beta, \hat{\varsigma})V_1^{-1}(\beta, \hat{\varsigma})D_1(\beta, \hat{\varsigma})\right\}^{-1} \\ &\times \left[D_1'(\beta, \hat{\varsigma})V_1^{-1}(\beta, \hat{\varsigma})\{y - \mu(\beta, \hat{\varsigma})\}\right]. \end{aligned} \tag{10}$$

The estimator $\hat{\beta}$ is consistent for β as the generalized estimation equation in (9) is an unbiased estimating equation. This estimator is also highly efficient (Liang & Zeger 1986). The reason is because $\hat{\beta}$ is obtained by solving the estimating equation (9), where the variance-covariance matrix $V_1(\beta, \varsigma)$ is the correct variance-covariance matrix of the responses. To show the asymptotic distribution of $\hat{\beta}$, we have the following Theorem.

THEOREM 1. *Under mild regularity conditions as $I \longrightarrow \infty$, $\hat{\beta}$ is a consistent estimator of β and that $I^{1/2}(\hat{\beta} - \beta)$ is asymptotically multivariate Gaussian with mean vector* 0 *and covariance matrix V_β given by*

$$V_\beta = I\{D_1'(\beta, \hat{\varsigma})V_1^{-1}(\beta, \hat{\varsigma})D_1(\beta, \hat{\varsigma})\}^{-1}. \quad (11)$$

The proof is based on standard approaches and is omitted for simplicity.

Note that the solution obtained from (9) requires ς to be known. In equation (9), $\hat{\varsigma}$ is treated as nuisance parameter. Similar to (9), we develop a GEE for ς to estimate the variance components ς.

3.2. Generalized estimating equations for random effects.

We now find the estimating equation with respect to variance components ς. This estimating equation can be written as

$$\sum_{i=1}^{I} D_{i2}'(\beta, \varsigma)V_{i2}^{-1}(\beta, \varsigma)\{S_i(\beta, \varsigma) - \sigma_i(\beta, \varsigma)\} = 0,$$

letting the independence assumption for $S_i(\beta, \varsigma) = (s_{i11}, ..., s_{i1T}, ..., s_{iT1}, ..., s_{iTT})'$ $\equiv (v_{i1}, v_{i2}, ..., v_{iT^2})'$ among areas, with $s_{ikl}(\beta, \varsigma) = (y_{ik} - \mu_{ik})(y_{il} - \mu_{il})$, and similarly $\sigma_i(\beta, \varsigma)$=$(\sigma_{i11}, \sigma_{i12}, ..., \sigma_{i1T},\ \sigma_{i21}, ..., \sigma_{i2T}$,...,$\sigma_{iT1}, ..., \sigma_{iTT})'$. With assuming the independence for $S_i(\beta, \varsigma)$ among areas, we may lose some efficiency. We study this issue through simulation study in Section 4.2. Moreover, $V_{i2}(\beta, \varsigma) = \text{cov}(S_i) = \{\tilde{\sigma}_{ijk}\}_{j,k=1}^{T^2}$, where $\tilde{\sigma}_{ikl} = cov(v_{ik}, v_{il}), i = 1, ..., I; k, l = 1, ..., T^2$. We also need to find $D_{i2}(\beta, \varsigma) = \partial\sigma_i(\beta, \varsigma)/\partial\varsigma$. To this end, we need to calculate $\partial\sigma_{ikl}/\partial\lambda_\eta$, $\partial\sigma_{ikl}/\partial\sigma^2_\eta, \partial\sigma_{ikl}/\partial\sigma^2_\alpha$, $\partial\sigma_{ikl}/\partial\rho$, and $\partial\sigma_{ikl}/\partial\sigma^2_\theta; k, l = 1, ..., T$. The derivation details of the components $V_{i2}(\beta, \varsigma)$ and $D_{i2}(\beta, \varsigma)$ are provided in the Appendix.

To estimate the variance components ς, we may obtain the GEE based estimate of ς by solving the estimating equation

$$\sum_{i=1}^{I} D'_{i2}(\hat{\beta}, \varsigma)V_{i2}^{-1}(\hat{\beta}, \varsigma)\{S_i(\hat{\beta}, \varsigma) - \sigma_i(\hat{\beta}, \varsigma)\} = 0. \tag{12}$$

Similar to equation (9), one may solve equation (12) for ς using the Newton-Raphson iterative method. Given the value $\hat{\varsigma}_s$ at the *s-th* iteration, $\hat{\varsigma}_{s+1}$ may be obtained at the *(s + 1)-th* iteration as

$$\begin{aligned} \hat{\varsigma}_{s+1} &= \hat{\varsigma}_s + \Big\{ \sum_{i=1}^{I} D'_{i2}(\hat{\beta}, \varsigma)V_{i2}^{-1}(\hat{\beta}, \varsigma)D_{i2}(\hat{\beta}, \varsigma)\Big\}^{-1} \\ &\times \Big[\sum_{i=1}^{I} D'_{i2}(\hat{\beta}, \varsigma)V_{i2}^{-1}(\hat{\beta}, \varsigma)\{S_i(\hat{\beta}, \varsigma) - \sigma_i(\hat{\beta}, \varsigma)\}\Big]. \end{aligned} \tag{13}$$

Note that in constructing the GEE for β, we have used the true covariance matrix $V_1(\beta, \varsigma)$, while the GEE for ς in (12) uses a *working* fourth-order matrix $V_{i2}(\beta, \varsigma)$. To show the asymptotic distribution of $\hat{\varsigma}$, we have the following Theorem.

THEOREM 2. *Under mild regularity conditions as $I \longrightarrow \infty$, $\hat{\varsigma}$ is a consistent estimator of ς and that $I^{1/2}(\hat{\varsigma} - \varsigma)$ is asymptotically multivariate Gaussian* *with mean vector* 0 *and covariance matrix* V_ς *given by*

$$\begin{aligned} V_\varsigma &= I\Big\{\sum_{i=1}^{I} D'_{i2}(\hat\beta,\varsigma)V_{i2}^{-1}(\hat\beta,\varsigma)D_{i2}(\hat\beta,\varsigma)\Big\}^{-1}\Big\{\sum_{i=1}^{I} D'_{i2}(\hat\beta,\varsigma)V_{i2}^{-1}(\hat\beta,\varsigma)F_i \\ &\times V_{i2}^{-1}(\hat\beta,\varsigma)D_{i2}(\hat\beta,\varsigma)\Big\}\Big\{\sum_{i=1}^{I} D'_{i2}(\hat\beta,\varsigma)V_{i2}^{-1}(\hat\beta,\varsigma)D_{i2}(\hat\beta,\varsigma)\Big\}^{-1}, \end{aligned} \tag{14}$$

where $F_i = E\{S_i(\hat\beta,\varsigma) - \sigma_i(\hat\beta,\varsigma)\}\{S_i(\hat\beta,\varsigma) - \sigma_i(\hat\beta,\varsigma)\}'$. The proof is based on standard approaches and is omitted for simplicity.

Note that F_i in Theorem 2 is replaced with $V_{i2}(\hat\beta,\varsigma)$ in our application (Section 4) and its misspecification is addressed in Section 4.2. The estimator $\hat\varsigma$ is consistent, but loses its efficiency slightly because of the use of a *working* covariance matrix in the estimating equation (12). The degree of loss of efficiency depends on the level of misspecification of the *working* fourth-order moments matrix to be used in the place of the true fourth-order moments matrix.

The complete algorithm for the GEE estimates of β and ς can be described as follows:

1. Choose initial values β_0 and ς_0. Set $m = 0$.
2. (a) Calculate β_{m+1} from the iterative equation (10).
   (b) Calculate ς_{m+1} from the iterative equation (13).
   (c) Set $m = m + 1$.
3. Continue step 2 until a convergence is achieved. Declare the estimates at convergence to be the GEE estimators $\hat\beta$ and $\hat\varsigma$ of β and ς.

4. APPLICATION

4.1. Data analysis.

We illustrate our model on presentations for asthma made by children (age < 18 years) to EDs in the province of Alberta, Canada. Presentations to the EDs and population data are provided by large, provincial databases. During the study period, Alberta had about 3.1 million population, of which around 0.8 million were children. We focus on the number of monthly pediatric presentations made between January 1, 2001, and December 31, 2004 ($T = 48$ months). The province consists of nine Regional Health Authorities that are responsible for the delivery of health care services. These nine regions are further sub-divided into ($I = 70$) sub-Regional Health Authorities (sRHAs) and these sRHAs are the geographic units used in our model. The visits totaled 62,008 over the study period with mean and median monthly visits per sRHA of 18 and 16 (range 0 to 111), respectively.

The design matrix $X = col_{1\leq i\leq I}(X_i)$ that corresponds to the fixed effects has dimension $N \times$ p, where $N = 3360$ and p $= 3$. The design matrix $Z = col_{1\leq i\leq I}(Z_i)$ that corresponds to the random effects has dimension $N \times$ h, where h $= 3478$.

Figure 1 presents the crude provincial asthma visits rate over period, which clearly shows the need of seasonal adjustment. We fit the model (1) to the data set of asthma visits (Table 1). Based on our findings on the structure of the main effects, we consider exchangeability of the components

TABLE 1: Parameter estimates (Est.) and standard errors (SE), spatio-temporal mixed model, monthly pediatric asthma visits to EDs, 2001-2004.

	Coefficients of the fixed effects			Variance components		
	Parameter	Est.	SE	Parameter	Est.	SE
Overall mean	μ	0.0004	0.0001	λ_η	0.10	0.06
				σ^2_η	3.06	0.52
Seasonal components	A	-0.13	0.05	σ^2_α	0.11	0.04
	B	-0.09	0.04	ρ	0.15	0.10
				σ^2_θ	0.05	0.01

of θ by taking $K_\theta = I$, the identity matrix (Knorr-Held 2000). Some model parameters are clearly significant (5% level) based on the asymptotic results given in Theorems 1 and 2 (equations (11) and (14)).

Formal tests for overdispersion may be used to indicate its presence in the model (1) (Dean & Lawless 1989; Dean 1992; Sinha 2009). We applied the parametric bootstrap approach proposed by Sinha (2009) to evaluate the significance of variance components and observed that there are overdispersion in both spatial and temporal effects.

To investigate diagnostic analysis, we calculate the *deviance residual* (McCullagh & Nelder 1989) as

$$dr_{it} = \text{sgn}(y_{it} - \hat{\mu}^c_{it})\left\{2\Big(y_{it}\log(\frac{y_{it}}{\hat{\mu}^c_{it}}) - y_{it} + \hat{\mu}^c_{it}\Big)\right\}^{1/2},$$

where

$$\text{sgn}(a) = \begin{cases} 1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}.$$

Note that we obtain the random terms based on $1,000$ Monte Carlo where

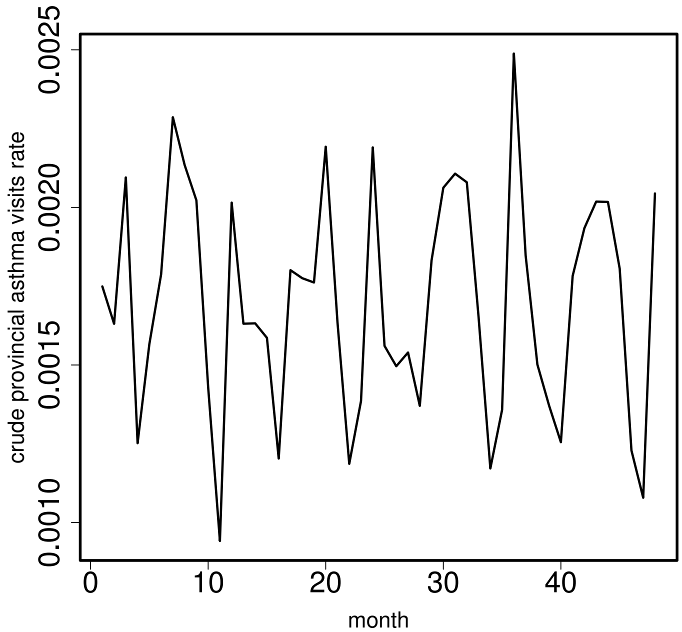

FIGURE 1: Crude provincial asthma visits rate over the period January 2001 to December 2004.

needed, for example $\hat{\mu}^c_{it}$. Figure 2 gives the residuals versus log-predicteds diagnostic plot. It is clear from Figure 2 that there is no serious lack of fit in our model.

For a close examination of the temporal change of the estimated spatial risk profile, we define the adjusted relative risk by $\exp(\eta_i + \theta_{it})$, which is automatically calibrated on a common base for the temporal effects (Knorr-Held 2000). The overall peak months are April, May, and September. We provide the estimated regional asthma visits ratio for April 2001-2004 where the months May and September had similar patterns. The estimated spatial effects based on the fitted model are presented in Figure 3. These maps suggest that sRHAs with relatively high numbers of asthma visits ratio are clustered in the south-central part of the province. Generally, the spatial

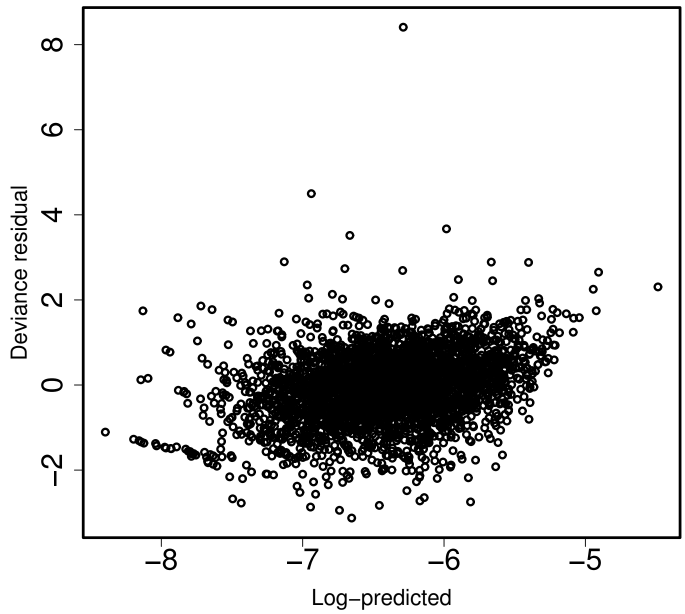

FIGURE 2: The deviance residuals versus log-predicteds diagnostic plot of pediatric asthma visits rate over the period January 2001 to December 2004.

pattern is relatively stable over time (Figure 4). Note that the regions with the highest ratios are generally in the rural regions. Individuals in these regions may not have as much access to alternative sources of health care in the non-ED setting or may have more severe disease.

We also provide the regional number of asthma visits ratio estimates obtained from fitting the spatio-temporal mixed model given by $\exp(\eta_i+\alpha_t+\theta_{it})$. Figure 5 plots the fitted asthma visits ratio for example for regions R312 and R617. The crude ratio estimates are $y_{it}/(n_{it}c_t)$, where $c_t = \sum_{i=1}^{I} y_{it}/\sum_{i=1}^{I} n_{it}$, and are also plotted in Figure 5. As shown, there are seasonal patterns over time in these regions. In general, a specific pattern in estimated log ratio over time for a region would suggest that the underlying asthma visits rate in that region has also the same pattern relative to the provincial average.

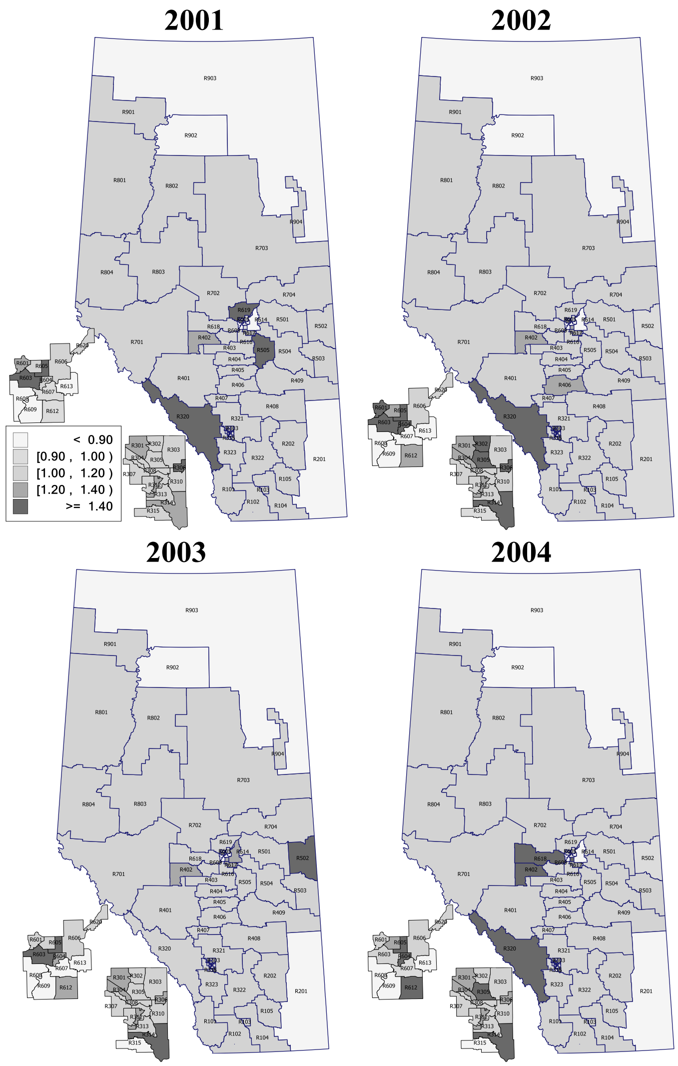

FIGURE 3: Adjusted relative risk for pediatric asthma visits to EDs in Alberta in April 2001-2004. Major urban centres are provided as inserts.

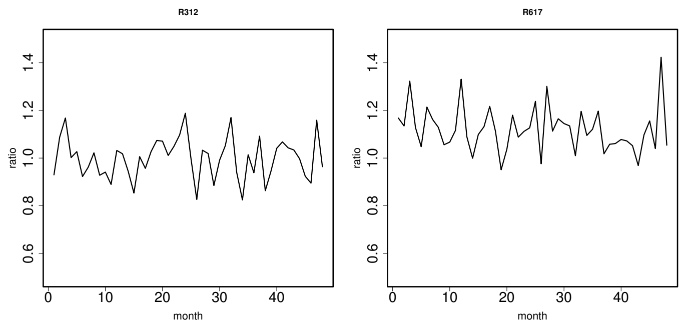

FIGURE 4: Adjusted relative risk for regions R312 and R617 from January 2001 to December 2004.

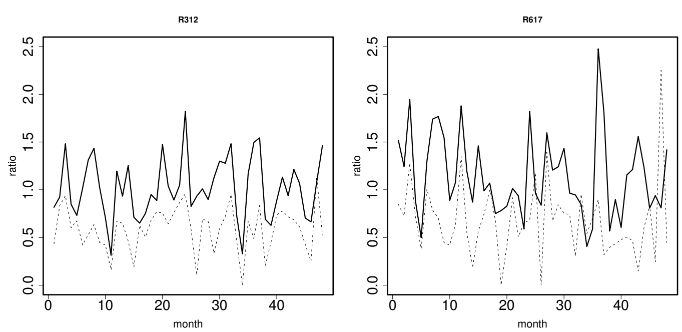

FIGURE 5: Fitted pediatric asthma visits ratio to EDs for selected regions, R312 and R617, January 2001 - December 2004. The solid line represents fitted ratios; the dashed line, crude ratios.

4.2. Simulation study on the performance of GEE estimates.

We conducted a simulation study to evaluate the performance of GEE estimates using the scenario similar to our asthma data. More specifically, data are generated from the model (1) with the parameters close to those obtained in the analysis of the asthma data; μ, A, B, and $\sigma_\eta^2, \lambda_\eta, \sigma_\alpha^2, \rho, \sigma_\theta^2$ are listed in Table 2. The neighborhood structure and the population sizes are exactly as for the asthma data. Estimates were obtained using GEE analyses of 500 data sets generated from the mixed Poisson model (1).

TABLE 2: Mean values of biases and standard errors, and simulated standard errors of GEE estimates based on 500 simulated data sets.

Parameter	Bias	Standard error $(X100)$	
		Simulated	GEE
$\mu = 0.0005$	-0.0061	0.015	0.013
$A = -0.15$	0.0032	5.6	5.3
$B = -0.10$	-0.0051	4.4	4.2
$\sigma_\eta^2 = 3.10$	0.0012	307.5	306.4
$\lambda_\eta = 0.10$	-0.0143	11.4	9.8
$\sigma_\alpha^2 = 0.10$	-0.0003	11.8	11.4
$\rho = 0.15$	0.0105	16.4	15.3
$\sigma_\theta^2 = 0.05$	-0.0001	5.4	4.8

Table 2 presents the bias values of the fixed parameters and variance component parameters, as well as the standard deviation of the estimated parameters and mean values of the estimated standard errors. The estimates are fairly unbiased, and it seems that their standard errors are estimated reasonably well. Overall, it seems that GEE provides good point estimates and standard errors for this data analysis. The use of the working covariance matrix to estimate the variance components is reasonable and provides a relatively small loss of information.

5. DISCUSSION

We have presented and illustrated a method for spatio-temporal analysis that pays specific attention to the mapping of area level disease rates over time. We have modeled the random spatial and temporal effects as well as the interaction between these two factors. Seasonal effects may be present and important to capture in some applications and our model incorporates those effects in a spatio-temporal model. The modeling approach is flexible and applicable to other contexts where spatio-temporal and seasonal effects exist.

For simplicity, we only considered the structure where all regions had the same seasonal effects. One can account for the effects of region i and time t, such that $A_{it}\cos(\pi t/6) + B_{it}\sin(\pi t/6)$ in model (1), where for example A_{it} and B_{it} are AR(1). Note that we also used $A_t\cos(\pi t/6) + B_t\sin(\pi t/6)$ where A_t and B_t were RW(1) and observed that there is no significant preference in using A_t and B_t rather than A and B. We did not include any covariate terms in the model. However, one can easily extend the model to include covariates and such inclusion may be required for some applications.

We used generalized estimating equations in this paper, although other estimation techniques could be used. Recently, Lele, Dennis & Lutscher (2007) proposed a method to compute the maximum likelihood estimation for hierarchical models using Bayesian MCMC methods. We have planned to study this approach in for our data.

Alternative analysis we have planned include a fully Bayesian approach using MCMC. More specifically, we will use highly vague, but proper, priors for all interested parameters to learn more from data. We will also investigate more complex models rather than (1), for example, A_{it} and B_{it}, instead of A and B, where A_{it} and B_{it} have AR(1) or RW(1), and study the sensitivity of such type of analysis to prior assumptions in the Poisson context.

APPENDIX

Derivation of the components of $V_{i2}(\beta, \varsigma)$. To calculate the components of $V_{i2}(\beta, \varsigma)$, we use the *independence* assumption for repeated y_{ij} within an area as a working *covariance matrix* (Sutradhar & Farrell 2004; Bari & Sutradhar 2005). We may lose some efficiency by assuming the independence for repeated y_{ij}. We address this issue through simulation study in Section 4.2. Hence, we have $\tilde{\sigma}_{i11} = \text{cov}(v_{i1}, v_{i1}) = E(y_{i1} - \mu_{i1})^4 =: \sigma^*_{i11}$,$\tilde{\sigma}_{i12} = \text{cov}(v_{i1}, v_{i2}) = E(y_{i1} - \mu_{i1})^3(y_{i2} - \mu_{i2}) = 0, ..., \tilde{\sigma}_{i1T} = \text{cov}(v_{i1}, v_{iT}) = 0$, and similarly $\tilde{\sigma}_{i1(T+2)} = \text{cov}(v_{i1}, v_{i(T+2)}) = E(y_{i1} - \mu_{i1})^2(y_{i2} - \mu_{i2})^2 = \sigma_{i11}\sigma_{i22}$, $\tilde{\sigma}_{i1(2T+3)} = \text{cov}(v_{i1}, v_{i(2T+3)}) = E(y_{i1} - \mu_{i1})^2(y_{i3} - \mu_{i3})^2 = \sigma_{i11}\sigma_{i33}$,... ,$\tilde{\sigma}_{i1T^2} = \text{cov}(v_{i1}, v_{iT^2}) = E(y_{i1} - \mu_{i1})^2(y_{iT} - \mu_{iT})^2 = \sigma_{i11}\sigma_{iTT}$, elsewhere 0. In a similar way, we can calculate other rows of $V_{i2}(\beta, \varsigma)$. Then, we need to calculate $\sigma^*_{ikk}, k = 1, ..., T$ where $\sigma^*_{i22} = \tilde{\sigma}_{i(T+2)(T+2)}, \sigma^*_{i33} = \tilde{\sigma}_{i(2T+3)(2T+3)}$ and so on. More precisely, to calculate $\sigma^*_{ikk} = E(y_{ik} - \mu_{ik})^4$, we need to find the fourth moment y_{ik}^4. We know that $y_{ik}|v \sim \text{Poisson}(\tilde{\mu}_{ik})$, where $\tilde{\mu}_{ik} = \exp(\log n_{ik} + x_{ik}'\beta + z_{ik}'v)$. Then, we have $E(y_{ik}|v) = \tilde{\mu}_{ik}$ and $\text{var}(y_{ik}|v) = \tilde{\mu}_{ik}$. On the other hand, we may write $E(y_{ik} - \mu_{ik})^4 = E(y_{ik}^4) - 4\mu_{ik}E(y_{ik}^3) + 6\mu_{ik}^2 E(y_{ik}^2) - 3\mu_{ik}^4$, where

$$E(y_{ik}^4) = EE(y_{ik}^4|v) = E(\tilde{\mu}_{ik}) + 7E(\tilde{\mu}_{ik}^2) + 6E(\tilde{\mu}_{ik}^3) + E(\tilde{\mu}_{ik}^4). \tag{15}$$

Recall that $E(\tilde{\mu}_{ik}) = \mu_{ik} = \exp\Big\{\log n_{ik} + x_{ik}'\beta + \frac{1}{2}\Big(\sigma_\theta^2 q_{ik}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1}\Big)\Big\}$. Similarly,

$$E(\tilde{\mu}_{ik}^2) = \exp\Big\{2\Big(\log n_{ik} + x_{ik}'\beta + \sigma_\theta^2 q_{ik}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1}\Big)\Big\},$$

$$E(\tilde{\mu}_{ik}^3) = \exp\Big[3\Big\{\log n_{ik} + x_{ik}'\beta + \frac{3}{2}(\sigma_\theta^2 q_{ik}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\Big\}\Big],$$

and

$$E(\tilde{\mu}_{ik}^4) = \exp\Big[4\Big\{\log n_{ik} + x_{ik}'\beta + 2(\sigma_\theta^2 q_{ik}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1})\Big\}\Big].$$

By combining the above results with (15), we obtain $E(y_{ik}^4)$. Moreover, we may write $E(y_{ik}^3) = EE(y_{ik}^3|v) = E(\tilde{\mu}_{ik}) + 3E(\tilde{\mu}_{ik}^2) + E(\tilde{\mu}_{ik}^3)$. Hence, $E(y_{ik}^3)$ is now known. In addition, we have $E(y_{ik}^2) = EE(y_{ik}^2|v) = E(\tilde{\mu}_{ik}) + E(\tilde{\mu}_{ik}^2)$, which is also known by above results. By combining the above results, we obtain $E(y_{ik} - \mu_{ik})^4$ after some simplification as

$$\begin{aligned}\sigma_{ikk}^* &= E(y_{ik} - \mu_{ik})^4 = \mu_{ik} - 4\mu_{ik}^2 + 6\mu_{ik}^3 - 3\mu_{ik}^4 \\ &\quad + (7 - 12\mu_{ik} + 6\mu_{ik}^2)\exp\Big\{2\Big(\log n_{ik} + x_{ik}'\beta + \sigma_\theta^2 q_{ik}^{-1} + \sigma_\alpha^2/(1-\rho^2) + \sigma_\eta^2 p_{ii}^{-1}\Big)\Big\}\end{aligned}$$

$$+(6-4\mu_{ik})\exp\Big[3\Big\{\log n_{ik}+x_{ik}^{'}\beta+\frac{3}{2}(\sigma_\theta^2 q_{ik}^{-1}+\sigma_\alpha^2/(1-\rho^2)+\sigma_\eta^2 p_{ii}^{-1})\Big\}\Big]$$
$$+\exp\Big[4\Big\{\log n_{ik}+x_{ik}^{'}\beta+2(\sigma_\theta^2 q_{ik}^{-1}+\sigma_\alpha^2/(1-\rho^2)+\sigma_\eta^2 p_{ii}^{-1})\Big\}\Big].$$

We use all of the results above to provide $V_{i2}(\beta,\varsigma)$.

Derivation of the components of $D_{i2}(\beta,\varsigma)$. To find $D_{i2}(\beta,\varsigma)$, we need to calculate $\partial\sigma_{ikl}/\partial\lambda_\eta, \partial\sigma_{ikl}/\partial\sigma_\eta^2, \partial\sigma_{ikl}/\partial\sigma_\alpha^2, \partial\sigma_{ikl}/\partial\rho$, and $\partial\sigma_{ikl}/\partial\sigma_\theta^2; k,l=1,...,T$. We have the following terms after some simplification

$$\partial\sigma_{ikk}/\partial\lambda_\eta=\mu_{ik}\sigma_\eta^2 h_{ii}\Big[2\mu_{ik}^3\exp\{-2(\log n_{ik}+x_{ik}^{'}\beta)\}-\mu_{ik}+1/2\Big];k=1,...,T,$$

$$\partial\sigma_{ikl}/\partial\lambda_\eta=-\sigma_\eta^2 h_{ii}\exp\Big\{\log n_{ik}+\log n_{il}+(x_{ik}+x_{il})'\beta\Big\}\Big[-2\exp\Big\{\sigma_\alpha^2\frac{1+\rho^{|k-l|}}{1-\rho^2}$$
$$+\sigma_\theta^2(q_{ik}^{-1}+q_{il}^{-1})+2\sigma_\eta^2 p_{ii}^{-1}\Big\}+\exp\Big\{\sigma_\theta^2(q_{ik}^{-1}+q_{il}^{-1})/2+\sigma_\alpha^2/(1-\rho^2)+\sigma_\eta^2 p_{ii}^{-1}\Big\}\Big];k\neq l=1,...,T,$$

where h_{ii} is the i-th diagonal element of $P^{-1}[I-D]P^{-1}$. In addition, we have

$$\partial\sigma_{ikk}/\partial\sigma_\eta^2=\mu_{ik}p_{ii}^{-1}\Big[2\mu_{ik}^3\exp\{-2(\log n_{ik}+x_{ik}^{'}\beta)\}-\mu_{ik}+1/2\Big];k=1,...,T,$$

$$\partial\sigma_{ikl}/\partial\sigma_\eta^2=\exp\Big\{\log n_{ik}+\log n_{il}+(x_{ik}+x_{il})'\beta\Big\}p_{ii}^{-1}\Big[2\exp\Big\{\sigma_\alpha^2\frac{1+\rho^{|k-l|}}{1-\rho^2}+\sigma_\theta^2(q_{ik}^{-1}+q_{il}^{-1})$$
$$+2\sigma_\eta^2 p_{ii}^{-1}\Big\}-\exp\Big\{\sigma_\theta^2(q_{ik}^{-1}+q_{il}^{-1})/2+\sigma_\alpha^2/(1-\rho^2)+\sigma_\eta^2 p_{ii}^{-1}\Big\}\Big];k\neq l=1,...,T.$$

In addition,

$$\partial\sigma_{ikk}/\partial\sigma_\alpha^2=\frac{\mu_{ik}}{1-\rho^2}\Big\{\mu_{ik}\Big[2\mu_{ik}^2\exp\{-2(\log n_{ik}+x_{ik}^{'}\beta)\}-1\Big]+1/2\Big\};k=1,...,T,$$

$$\partial\sigma_{ikl}/\partial\sigma^2_\alpha = \exp\Big\{\log n_{ik}+\log n_{il}+(x_{ik}+x_{il})'\beta\Big\}(1-\rho^2)^{-1}\Big[(1+\rho^{|k-l|})\exp\Big\{\sigma^2_\alpha\frac{1+\rho^{|k-l|}}{1-\rho^2}$$

$$+\sigma^2_\theta(q^{-1}_{ik}+q^{-1}_{il})+2\sigma^2_\eta p^{-1}_{ii}\Big\}-\exp\Big\{\sigma^2_\theta(q^{-1}_{ik}+q^{-1}_{il})/2+\sigma^2_\alpha/(1-\rho^2)+\sigma^2_\eta p^{-1}_{ii}\Big\}\Big]; k\neq l=1,...,T.$$

Moreover,

$$\partial\sigma_{ikk}/\partial\rho = \frac{2\rho\sigma^2_\alpha}{(1-\rho^2)^2}\mu_{ik}\Big[2\mu^3_{ik}\exp\{-2(\log n_{ik}+x'_{ik}\beta)\}-\mu_{ik}+1/2\Big]; k=1,...,T,$$

and for $k\neq l=1,...,T$,

$$\partial\sigma_{ikl}/\partial\rho = \frac{2\rho\sigma^2_\alpha}{(1-\rho^2)^2}\exp\Big\{\log n_{ik}+\log n_{il}+(x_{ik}+x_{il})'\beta\Big\}\Big[\Big\{1+\rho^{|k-l|}+|k-l|(1-\rho^2)\rho^{|k-l|-2}/2\Big\}$$

$$\times\exp\Big\{\sigma^2_\alpha\frac{1+\rho^{|k-l|}}{1-\rho^2}+\sigma^2_\theta(q^{-1}_{ik}+q^{-1}_{il})+2\sigma^2_\eta p^{-1}_{ii}\Big\}-\exp\Big\{\sigma^2_\theta(q^{-1}_{ik}+q^{-1}_{il})/2+\sigma^2_\alpha/(1-\rho^2)+\sigma^2_\eta p^{-1}_{ii}\Big\}\Big].$$

Furthermore,

$$\partial\sigma_{ikk}/\partial\sigma^2_\theta = q^{-1}_{ik}\mu_{ik}\Big[2\mu^3_{ik}\exp\{-2(\log n_{ik}+x'_{ik}\beta)\}-\mu_{ik}+1/2\Big]; k=1,...,T,$$

$$\partial\sigma_{ikl}/\partial\sigma^2_\theta = (q^{-1}_{ik}+q^{-1}_{il})\exp\Big\{\log n_{ik}+\log n_{il}+(x_{ik}+x_{il})'\beta\Big\}\Big[\exp\Big\{\sigma^2_\alpha\frac{1+\rho^{|k-l|}}{1-\rho^2}+\sigma^2_\theta(q^{-1}_{ik}$$

$$+q^{-1}_{il})+2\sigma^2_\eta p^{-1}_{ii}\Big\}-\frac{1}{2}\exp\Big\{\sigma^2_\theta(q^{-1}_{ik}+q^{-1}_{il})/2+\sigma^2_\alpha/(1-\rho^2)+\sigma^2_\eta p^{-1}_{ii}\Big\}\Big]; k\neq l=1,...,T.$$

We obtain $D_{i2}(\beta,\varsigma)$ by combining the above results.

REFERENCES

W. Bari & B.C. Sutradhar (2005). On binary longitudinal mixed models in adaptive clinical trials. *Journal of Statistical Computation and Simulation,* 75, 687-707.

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

J. Besag, J. York & A. Molliè (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics,* 43, 1-21.

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

D.G. Clayton (1996). Generalized linear mixed models, *In Markov Chain Monte Carlo in Practice,* Ed. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, pp. 275-301. London: Chapman and Hall.

D.G. Clayton & L. Bernardinelli (1992). Bayesian methods for mapping disease risk. *In Geographical and Environmental Epidemiology: Methods for Small-Area Studies,* Ed. P. Elliott, J. Cuzick, D. English, and R. Stern, pp. 205-220. Oxford: Oxford University Press.

D.G. Clayton & J. Kaldor (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics,* 43, 671-681.

G.S. Datta, P. Lahiri, T. Maiti & K.L. Lu (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association,* 94, 1074-1082.

C.B. Dean (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association,* 87, 451-457.

C.B. Dean & J.F. Lawless (1989). Tests for detecting overdispersion in Poisson regresion models. *Journal of the American Statistical Association,* 84, 467-472.

P.H.C. Eilers & B.D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science,* 11(2), 89-121.

T.R. Fanshawe, P.J. Diggle, S. Rushton, P.W.W. Lurz, S.V. Glinianaia, M.S. Pearce, L. Parker, M. Chalton & T. Pless-Mulloli (2008). Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics,* 19, 549-566.

W.R. Gilks, S. Richardson & D.J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice.* New York: Springer -Verlag.

C.Gössl, D.P. Auer & L. Fahrmeir (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics,* 57, 554-562.

S.K. Greene, E.L. Ionides & M.L. Wilson (2006). Patterns of influenza-associated mortality among U.S. elderly by geographic region and virus subtype, 1968-1998. *American Journal of Epidemiology,* 163 (4), 316-326.

D.A. Harville (1997). *Matrix Algebra from a Statistician's Perspective.* Springer, New York.

V. Jowaheer & B.C. Sutradhar (2002). Analyzing longitudinal count data with over dispersion. *Biometrika,* 89, 389-399.

L. Knorr-Held (2000). Bayesian modeling of inseparable space-time variation in disease risk. *Statistics in Medicine,* 19, 2555-2567.

L. Knorr-Held & S. Richardson (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics,* 52(2), 169-183.

A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel & R. Bertollini

(1999). *Disease Mapping and Risk Assessment for Public Health*. West Sussex, U.K.: John Wiley and Sons Ltd.

Y. Lee & J.A. Nelder (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B,* 58, 619-678.

S.R. Lele, B. Dennis & F. Lutscher (2007). Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters,* 10, 551-563.

B.G. Leroux, X. Lei & N. Breslow (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *In Statistical Models in Epidemiology, the Environment and Clinical Trials,* Ed. M.E. Halloran and D. Berry, pp. 135-178. New York: Springer-Verlag.

K.Y. Liang & S.L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 73, 13-22.

X. Lin & N.E. Breslow (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association,* 91, 1007-1016.

M.L.H. Mabaso, M. Craig, P. Vounatsou & T. Smith (2005). Towards empirical distribution of malaria seasonality in southern Africa: the example of Zimbabwe. *Tropical Medicine and International Health,* 10 (9), 909-918.

Y.C. MacNab & C.B. Dean (2000). Parametric bootstrap and penalized quasi-likelihood inference in conditional autoregressive models. *Statistics in Medicine,* 19 (17/18), 2421-2435.

Y.C. MacNab & C.B. Dean (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics,* 57, 949-956.

M.A. Martínez-Bereito, A. López-Quilez & P. Botella-Rocamora (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine,* 27, 2874-2889.

P. Mccullagh (1983). Quasi-likelihood functions. *Annals of Statistics,* 11, 59-67.

P. Mccullagh & J.A. Nelder (1989). *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

C.E. Mcculloch (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association,* 92, 162-190.

R.L. Prentice & L.P. Zhao (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics,* 47, 825-839.

G.L. Silva, C.B. Dean, T. Niyonsenga & A. Vanasse. Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine,* 27, 2381-2401.

S.K. Sinha (2009). Bootstrap tests for variance components in generalized linear mixed models. *The Canadian Journal of Statistics,* 37, 219-234.

B.C. Sutradhar & K. Das (1999). On the efficiency of regression estimation in generalized linear models for longitudinal data. *Biometrika,* 86, 459-465.

B.C. Sutradhar & P.J. Farrell (2004). Analyzing multivariate longitudinal binary data: A generalized estimating equation approach. *Canadian Journal of Statistics,* 32, 39-55.

J. Wakefield & S. Morris (1999). Spatial dependence and error-in-variables in environmental epidemiology. *In Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. Smith, pp. 657-684. Oxford: Clarendon Press.

L.A. Waller, B.P. Carlin, H. Xia & A.E. Gelfand (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association,* 92, 607-617.

R. Wedderburn (1974). Quasi-likelihood functions, generalized linear models, and gauss-newton method. *Biometrika,* 61, 439-447.

Y. Yasui & S. Lele (1997). A regression method for spatial disease rates: An estimating function approach. *Journal of the American Statistical Association,* 92, 21-32.

S.L. Zeger (1988). A regression model for time series of counts. *Biometrika,* 75, 621-629.

S.L. Zeger & K.Y. Liang (1986). Longitudinal data analysis of discrete and continuous outcomes. *Biometrics,* 42, 121-130.