Spatial modeling of repeated events with an application to disease mapping

Debarghya Jana Niranjan Dey
Rishiraj Sutar Shatrughna Chaurasia

MTH643A Presentation

Supervisor:
Dr. Arnab Hazra

Department Of Mathematics and Statistics

Indian Institute Of Technology, Kanpur

Date: November 14, 2023

Presentation Contents

Introduction

- Mixed models are commonly used to analyze spatial data which frequently occur in practice such as in health sciences and life studies.
- It is important to incorporate spatial random effects into the model to account for spatial variation of the data.
- Generally, Poisson mixed models are used to analyze spatial count data.
- It is often assumed that the observations in each area, conditional on the spatial random effects, are independent of each other.
- But, this is not a valid assumption in practice.
- In particular, compound Poisson mixed models are introduced to account for the repeated events as well as the spatial variation of the data.
- To address this issue, we implement 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.

Spatial compound Poisson model

- Let C_{iw} be the random variable representing the number of individuals with exactly w events at area $i(=1,\ldots,m)$, 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 j th individual $(j = 1,\ldots,C_i)$ at area i. Hence, $Y_i = \sum_w wC_{iw} = \sum_{j=1}^{C_i} F_{ij}$ is the random variable representing the total number of events at area i.
- 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 $\lambda_w\lambda_i\left(1+\lambda_w\right)$. In particular, we can write

$$\lambda_i = \exp\left(\log e_i + \mathbf{x}_i^\top\boldsymbol{\beta} + \mathbf{z}_i^\top\boldsymbol{\eta}\right) \quad (1)$$

Spatial compound Poisson model(Continued)

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

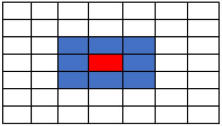
$$\begin{aligned} &\boldsymbol{\eta} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right) \\ &\boldsymbol{\Sigma}_\eta = \sigma_\eta^2\left[\left(1-\lambda_\eta\right)\mathbf{I_m} + \lambda_\eta \mathbf{R}\right]^{-1} \end{aligned} \tag{2}$$

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

- Here $I\{i \sim j\}$ defined as follows,

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

  where N_i is the set of blue regions.



Full Model set up

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

$$g[E(\boldsymbol{Y} \mid \boldsymbol{\eta})] = \text{ offset } + \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\eta}, \tag{3}$$

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

Naive Model: Spatial Poisson model

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

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

with mean μ_i given by

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

- Here, $log(e_i)$ denotes the offset for ith region. The covariate vector for the i-th observation including an intercept is given by $\mathbf{x}_i = (1, x_{i1}, .., x_{ip})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, .., \beta_p)'$ denotes the vector of unknown regression parameters. Spatial dependencies are modelled by introducing a random effect $\gamma_i, i = 1, ., m$ for each region.

Naive Model: Spatial Poisson model

- $\mathbf{z}_i^\top(m \times 1)$ is a known design vector, and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^\top$ represent spatial random effects.
- We consider the following general model for the spatial random effects η :

$$\begin{aligned} \boldsymbol{\eta} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\ \boldsymbol{\Sigma}_\eta &= \sigma_\eta^2 \left[(1-\lambda_\eta)\mathbf{I_m} + \lambda_\eta \mathbf{R}\right]^{-1} \end{aligned} \tag{4}$$

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

Quasi-likelihood approach

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

$$\sigma_{ii}(\boldsymbol{\theta}) \equiv \operatorname{Var}\left(Y_i\right) = \mu_i \left\{\exp\left(\log e_i + \mathbf{x}_i^\top \boldsymbol{\beta}\right) \lambda_w \left[\exp\left(\frac{3}{2}\Sigma_\eta^{ii}\right) - \exp\left(\frac{1}{2}\Sigma_\eta^{ii}\right)\right] + (1 + \lambda_w)\right\}, \tag{5}$$

and,

$$\sigma_{ij}(\boldsymbol{\theta}) \equiv \operatorname{cov}\left(Y_i, Y_j\right) = \lambda_w^2 \exp\left[\log e_i + \log e_j + \left(\mathbf{x}_i + \mathbf{x}_j\right)^\top \beta\right] \left\{\exp\left[\frac{1}{2}\left(\Sigma_\eta^{ii} + \Sigma_\eta^{ij}\right)\right] \times \left[\exp\left(\Sigma_\eta^{ij}\right) - 1\right]\right\} \tag{6}$$

Quasi-likelihood equations for fixed effects

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

$$\mathbf{D}_1^\top\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2\right)\right] = \mathbf{0} \tag{7}$$

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

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1^{(k+1)} \approx \hat{\boldsymbol{\theta}}_1^{(k)} &+ \left[\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \\ &\times \left\{\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]\right\}.\end{aligned} \tag{8}$$

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

$$\widehat{\operatorname{var}\left(\hat{\boldsymbol{\theta}}_1\right)} \approx \left[\mathbf{D}_1^\top(\hat{\boldsymbol{\theta}})\mathbf{V}_1^{-1}(\hat{\boldsymbol{\theta}})\mathbf{D}_1(\hat{\boldsymbol{\theta}})\right]^{-1} \tag{9}$$

Quasi-likelihood equations for variance components

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

$$\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\left[\boldsymbol{S}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)\right] = \mathbf{0}, \qquad (10)$$

- Here $\mathbf{D}_2(\boldsymbol{\theta}) = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_2$ and $\hat{\boldsymbol{\theta}}_1$ is the estimation of $\boldsymbol{\theta}_1$.
- We can then use, for example, the Newton-Raphson iterative to estimate the variance components $\boldsymbol{\theta}_2$ from the QL estimating equations (6). In particular, given the value $\hat{\boldsymbol{\theta}}_2^{(k)}$ at the k th iteration, $\hat{\boldsymbol{\theta}}_2^{(k+1)}$ is obtained at the $(k+1)$ th iteration as:

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

Algorithm

Algorithm 1 Complete algorithm to estimate the model parameters

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

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

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1^{(k+1)} \approx \hat{\boldsymbol{\theta}}_1^{(k)} &+ \left[\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{D}_1\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]^{-1} \\ &\times\left\{\mathbf{D}_1^\top\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\mathbf{V}_1^{-1}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\left[\boldsymbol{Y} - \boldsymbol{\mu}\left(\hat{\boldsymbol{\theta}}_1^{(k)}, \hat{\boldsymbol{\theta}}_2\right)\right]\right\}.\end{aligned} \tag{13}$$

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

$$\begin{aligned}\hat{\boldsymbol{\theta}}_2^{(k+1)} \approx \hat{\boldsymbol{\theta}}_2^{(k)} &+ \left[\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{D}_2\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]^{-1} \\ &\times\left\{\mathbf{D}_2^\top\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\mathbf{V}_2^{-1}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\left[\boldsymbol{S}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right) - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2^{(k)}\right)\right]\right\}\end{aligned} \tag{14}$$

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

Disease ratio

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

$$SDR_i = \frac{\hat{\lambda}_w \hat{\lambda}_i}{e_i} = \hat{\lambda}_w \exp\left(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}_i\right) \tag{15}$$

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

$$\hat{\eta}_i = E\,\widehat{(\eta_i \mid Y_i)} = E\left(\eta_i \mid Y_i\right)\big|_{\theta=\hat{\theta}} \tag{16}$$

Data Application

- We have used a dataset of children (age < 18 years) asthma visits to physicians in the Canadian provinces of Manitoba & Winnipeg during the 2000–2009 to apply our proposed model.
- The population of these states were stable during the study period from 2000 to 2009 (1.15 million to 1.22 million). Thus we can compare this data over these years.
- Now, the State of Manitoba and Winnipeg consists of 8 regional health authoritis (7 in Manitoba and 1 in winnipeg).
- These states consist of 8 regional health authorities. These 8 regions were further subdivided into 67 Regional Health Authorities Districts (RHAD).
- The RHAD are the geographic units (areas) used in our model $i = 1(1)67$.

Data Application

- The dataset consists of 11825 rows and 5 columns where the first column indicates the spatial region, the second column indicates the year, third column indicates sex. The fourth column indicates the number of repetitions or w and the last column indicates the number of individuals with 'w' number of repetitions.
- The total number of asthma visits in the ith location denoted by Y_i will be calculated by multiplying the column 4 and column 5 and then summing the values w.r.t. i^{th} location and it is the variable of interest and acts as the response of the proposed model.
- The number of children's asthma visits totalled 694,484 over the years 2000-2009.

Data Application

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

Figure 1: w is the number of repeated events and C_w is the number of visits

Data Application

- It is clear from the frequency table that we have a range of 1 to 68 number of visits as repeated events and we can see the number of individuals is only 1 for more than 41 repeated events.

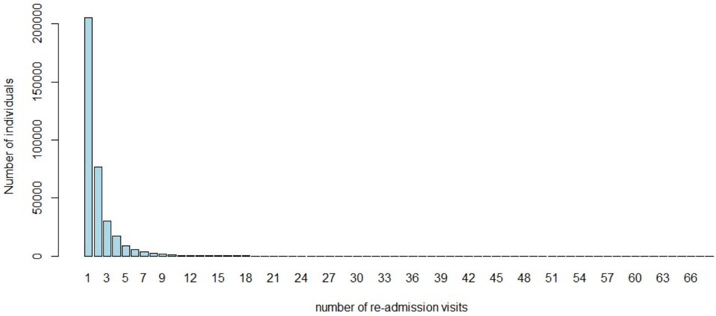


Figure 2: Barplot of Total number of asthma visits in i^{th} location (Y_i)

Data Application

- The dependency of total asthma visits in each region can be observed from the following barplots of Y_i and e_i.

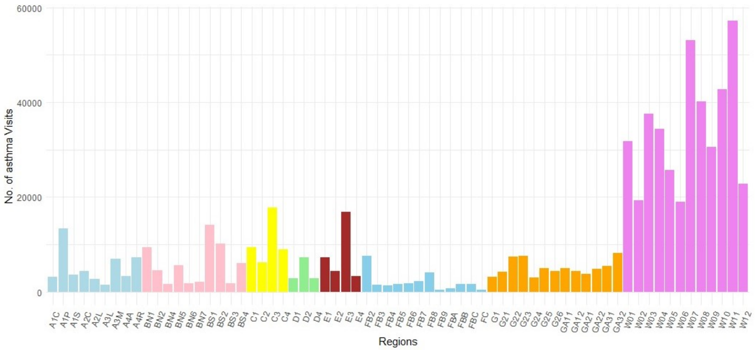


Figure 3: Barplot of Total number of asthma visits in i^{th} location (Y_i)

Data Application

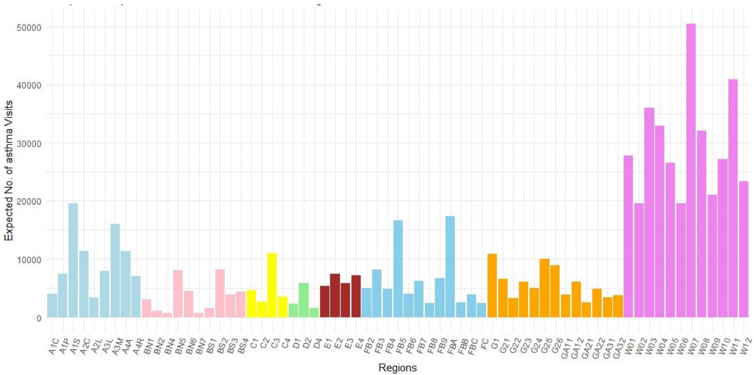


Figure 4:Barplot of expected number of asthma visits in i^{th} location (e_i)

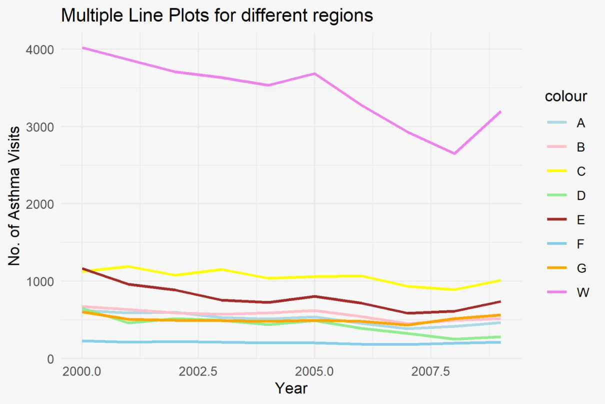


Figure 5:Line diagram of total no. of asthma visits for different years

- We can see, that the number of asthma visits in the province of Winnipeg is considerably larger than in the province of Manitoba and the asthma visits are similar within the provinces.
- Thus, from the above three plots, we can conclude that the repeated events, i.e., the number of asthma visits are not independent and spatial dependency is present.

Simulation study

- We conduct simulation studies to evaluate parameters and compare performance of our proposed spatial compound Poisson approach with the naive spatial Poisson approach.
- The Naive approach ignores the dependence of observations in each area given.
- 1200 datasets are generated from the both proposed and naive models.
- Data are generated from the model (1) with the true parameters equal to $\beta_0 = -2.643$, $\lambda_w = 5.565$, $\lambda_\eta = 0.478$ and,$\sigma_\eta = 1.959$ as specified in the paper.
- We used the Newton-Raphson algorithm to estimate the model parameters with a maximum of 500 iterations at each step and, the convergence is checked using a tolerance of 0.009.

Simulation study

- Here we use the spatial layout of our Manitoba RHADs, which is an irregular neighbourhood structure. The spatial structure is given in the data.
- The neighbourhood structure and the expected counts are given in the asthma visit dataset.

Parameter	True value	Proposed model		Naive model	
		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

Table 1:Bias and mean squared error (MSE) of the model parameter estimates for the proposed model (spatial compound Poisson model) and naive model (spatial Poisson model) in the case of an irregular grid.

- Table 2 presents the bias and MSE of the model parameter estimates for the proposed and naive models.

Simulation study

- We provide boxplots of model parameter estimates for both proposed and naive models to have a better understanding of the performance of these models.
- It appears that biases and MSEs of the model parameter estimates for our proposed model is considerably smaller than the corresponding values from the naive model.
- We set the parameter values similar to the values obtained in the data application ($\beta_0 = -2.643$, $\lambda\eta = 0.5$, and $\sigma_\eta = 2$) with four different values for λ_w as $1, 5, 10$, and 15 to evaluate the efficiency of the proposed model with the naive model
- The number of events per individual increases as λ_w is the mean of repeated events and we check for increasing values of λ_w.
- The biases and MSEs of model parameter estimates in the case of proposed model are consistently smaller than the values in the case of naive model.

Simulation study

Model	True parameter λ_w	β_0		λ_η		σ_η		λ_w	
		-2.5		0.5		2			
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Proposed	1	-0.207	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	-	-

Table 2:Bias and mean squared error (MSE) of the model parameter estimates for the proposed model (spatial compound Poisson model) and naive model (spatial Poisson model) for different values of λ_w in the case of an irregular grid.

Simulation study

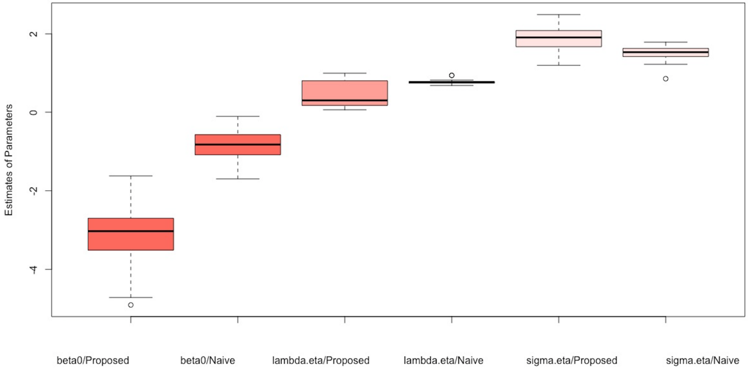


Figure 6:Box plots of the model parameter estimates for the proposed and naive models in the case of irregular grid

Conclusion

- In this paper, we implemented 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 conditional auto-regressive model for the spatial random effects.
- We implemented the quasi-likelihood method to estimate the model parameters.
- Our simulation studies also clearly showed the out-performance 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 the 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.
- **Future work:** Another interesting contribution will be to extend this approach to binary data which has many applications (Torabi, 2014). These are some of the topics for future study.

Appendix

Appendix 1: We will derive the link function as $\{\log(\cdot) - \log \lambda_w\}$. We can write

$$\log(\lambda_i) = \text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta} \tag{17}$$

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

$$E(Y_i \mid \boldsymbol{\eta}) = \lambda_w \exp\left(\text{ offset } + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\eta}\right) \tag{18}$$

Appendix(Continued)

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

$$\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} = \mathbf{x}_i^\top \left[\lambda_w \exp\left(\log e_i + \mathbf{x}_i^\top\boldsymbol{\beta}\right) M_\eta\left(\mathbf{z}_i\right)\right] = \mathbf{x}_i^\top\mu_i, \tag{19}$$

and

$$\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\lambda_w} = \exp\left(\log e_i + \mathbf{x}_i^\top\boldsymbol{\beta}\right) M_\eta\left(\mathbf{z}_i\right) = \mu_i/\lambda_w \tag{20}$$

References

[1] Leroux B.G., Lei X., Breslow N.(2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence: Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer (2000), pp. 179-191

[2] Susanne Gschlößl and Claudia Czado (2007)Spatial modelling of claim frequency and claim size in non-life insurance: Scandinavian Actuarial Journal

[3] Torabi M.(2014) Spatiotemporal modelling of odds of disease: Environmetrics,(2014), pp. 341-350

[4] Shabnam Balamchi and Mahmoud Torabi(2020): Spatial modelling of repeated events with an application to disease mapping: Spatial Statistics(2020)

Thank you