Biometrical Journal **58** (2016) 6, zzz–zzz / DOI: 10.1002/bimj.200100000

Zero-inflated spatio-temporal models for disease mapping

Mahmoud Torabi[*,1]

[1] *Department of Community Health Sciences, University of Manitoba, S113 Medical Services Building, 750 Bannatyne Ave., Winnipeg, MB, Canada R3E 0W3.*

Received zzz, revised zzz, accepted zzz

In this paper, our aim is to analyze geographical and temporal variability of disease incidence when spatio-temporal count data have excess zeros. To that end, we consider random effects in zero-inflated Poisson models to investigate geographical and temporal patterns of disease incidence. Spatio-temporal models that employ conditionally autoregressive smoothing across the spatial dimension and B-spline smoothing over the temporal dimension are proposed. The analysis of these complex models is computationally difficult from the frequentist perspective. On the other hand, the advent of the Markov chain Monte Carlo algorithm has made the Bayesian analysis of complex models computationally convenient. Recently developed data cloning method provides a frequentist approach to mixed models which is also computationally convenient. We propose to use data cloning, which yields to maximum likelihood estimation, to conduct frequentist analysis of zero-inflated spatio-temporal modeling of disease incidence. One of the advantages of the data cloning approach is that the prediction and corresponding standard errors (or prediction intervals) of smoothing disease incidence over space and time is easily obtained. We illustrate our approach using a real dataset of monthly children asthma visits to hospital in the province of Manitoba, Canada, during the period April 2006 to March 2010. Performance of our approach is also evaluated through a simulation study.

Key words: Bayesian computation; Hierarchical models; Random effects; Spatial models; Spline; Zero inflated models

1 Introduction

The analysis of disease incidence (or mortality) over space and time has received considerable attention due to growing demand for reliable disease mapping. The idea behind developments on spatio-temporal modeling of disease incidence is essentially to model variation in observed disease patterns and better separate systematic variability from random noise, a component that usually overshadows crude disease incidence maps. Maps of areal disease incidence over time are useful tools in determining spatial and temporal patterns of disease incidence for targeting resources. Disease incidence rates (ratios) may differ substantially across geographical areas. A reliable estimate of the underlying disease risk is usually provided by *borrowing strength* from neighboring 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 (1987) extended the use of mixed models for geographical data to account for the extra-Poisson variability through the introduction of spatial random effects in the context of disease mapping. However, there are many situations that our count data have excess zeros. A zero-inflated Poisson (ZIP) regression model was used to account for count data with excess zeros (Lambert, 1992). Following Lambert (1992), many works have been done in the case of count

[*]Corresponding author: e-mail: Mahmoud.Torabi@umanitoba.ca, Phone: +1-204-272-3136, Fax: +1-204-789-3905

data with excess zeros (Welsh *et al.*, 1996; Böhning, 1998; Yau *et al.*, 2003; Tse *et al.*, 2009; Wan and Chan, 2009). In the case of spatial count data with excess zeros, spatial ZIP models were employed (Agarwal *et al.*, 2002; Ugarte *et al.*, 2004; Rathbun and Fei, 2006). Recently, Nieto-Barajas and Bandyopadhyay (2013) considered a zero-inflated spatial gamma process model with applications to disease mapping. The literature is also available for zero-inflated models with continuous responses where a recent special issue of Biometrical Journal (Vol. 58 (2), 2016) was devoted on this subject.

The temporal random smoothing of incidence cases has also been studied in the literature. An autoregressive (AR) model for temporal count data was used by Zeger (1988). Waller *et al.* (1997) extended the hierarchical Bayesian spatial models to account for temporal random effects and spatio-temporal interactions. A unified approach for a Bayesian analysis of disease incidence in space and time was proposed by Knorr-Held (2000). MacNab and Dean (2001), Silva *et al.* (2008), and Torabi and Rosychuk (2012) proposed spatio-temporal models that use AR local smoothing across the spatial dimensions and B-spline smoothing over the temporal dimensions. Martinez-Beneito *et al.* (2008) suggested an AR spatio-temporal model in a Bayesian framework to link information in time and space. In some contexts, the underlying disease risks may change over seasons within a given year. Torabi and Rosychuk (2010) proposed a spatio-temporal model that uses conditional AR (CAR) smoothing across the spatial effects, AR smoothing over the temporal effects, and a smoothing function to account for seasonal effects. Torabi (2012) proposed a spatio-temporal model that uses AR smoothing across the spatial effects, random walk smoothing over the temporal effects, and a smoothing function to account for seasonal effects.

In some situations, spatio-temporal count data may have excess zeros. Wikle and Anderson (2003) considered a ZIP spatio-temporal model in the regression form where the covariates also depend on time, however, they did not consider the spatial and temporal random effects separately. Hoef and Jansen (2007) studied a ZIP spatio-temporal by having first-order AR model for the temporal random effects and CAR models for the spatial random effects. The both above papers used the Bayesian approach for the inference. There are many different ways to perform inference in mixed models, however, the frequentist approach has been computationally difficult particularly for these kinds of ZIP spatio-temporal models which fall in the class of generalized linear mixed models (GLMMs). Consequently, many approximate approaches have been proposed in the last two decades such as generalized estimating equations (Liang and Zeger, 1986; Prentice and Zhao, 1991; Torabi and Rosychuk, 2010) and penalized quasi-likelihood (Breslow and Clayton, 1993; Torabi and Rosychuk, 2011) among other approaches. With advances in computational power, the Bayesian approach especially the non-informative Bayesian approach has become quite popular although the implementation of the non-informative Bayesian approach requires substantial care.

Recently, Lele *et al.* (2007) introduced an alternative frequentist approach, called data cloning (DC), to compute the maximum likelihood (ML) estimates (MLE) and their standard errors for general hierarchical models. Similar to the Bayesian approach, the DC avoids high dimensional numerical integration and requires neither maximization nor differentiation of a function. Extending this work to GLMM situation, Lele *et al.* (2010) described an approach to compute prediction and prediction intervals for the random effects. The DC approach, thus, is well suited to offer a frequentist analysis of the zero-inflated spatio-temporal models.

The contribution of this paper is two-fold. First, we propose a zero-inflated spatio-temporal model to account for spatial and temporal count data with excess zeros. Second, we propose to use the DC as a frequentist approach for the inference. In the next section, we describe the zero-inflated spatio-temporal model. We then describe how the DC can be used to obtain model parameters estimate and predictions with corresponding standard errors (or prediction intervals) for smoothing disease incidence over space and time (Section 3). In Section 4, performance of the proposed approach is evaluated using a real dataset of monthly number of children asthma visits to hospital in the province of Manitoba, Canada, during the period April 2006 to March 2010. Performance of the ML estimate is also studied through a simulation study. Concluding remarks are given in Section 5. The Appendix is devoted to the proof of asymptotic normality of model parameters estimate.

2 Zero-inflated spatio-temporal model

Let y_{it} be the number of disease incidence cases (or otherwise) for the *i*-th geographic area at time t, and let e_{it} be the corresponding expected number of disease incidence cases for $i = 1, ..., I$; $t = 1, ..., T$. We propose the following zero-inflated spatio-temporal model:

$$y_{it}|(\mu_{it}, \theta_{it}) = \begin{cases} 0 & \text{with probability} \quad 1 - \theta_{it} \\ \text{Poi}(\mu_{it}) & \text{with probability} \quad \theta_{it} \end{cases} \tag{1}$$

where $\text{Poi}(\mu_{it})$ is a conditionally independent Poisson variable with mean function μ_{it} for the *i-th* geographic area at time t. In particular, we use link functions to relate the means of these distributions to linear mixed models,

$$\begin{aligned} &\log(\mu_{it}) = \log(e_{it}) + m + S(t) + \eta_i + \delta_{it}, \\ &logit(\theta_{it}) = m_0 + S_0(t) + \eta_{0i} + \delta_{0it}, \end{aligned} \tag{2}$$

which are combinations of fixed $(m, m_0, S(t), S_0(t))$ and random effects $(\eta_i, \delta_{it}, \eta_{0i}, \delta_{0it})$, where m is a fixed effect representing the overall mean log-ratio over time and area. To account for the fixed temporal effects, $S(t)$ represents a B-spline model with appropriate degree (order) and number of inner knots (Eilers and Marx, 1996). For instance, a cubic B-spline with three inner knots was found useful in our exploration of the data (Section 4). For our cubic B-spline, the knots are located at the first and third quartiles as well as at the median of time $t(= 1, ..., T)$; one can similarly define the number and location of inner knots for other (degrees of) B-spline models. To show the advancements of our proposed model, we focus on the cubic B-spline in the rest of the paper. With the overall mean of log-ratio m in our model, the B-spline is provided without an intercept. Hence, in the case of cubic B-spline model, $S(t)$ is given by

$$S(t) = \beta_1 B_1(t) + \beta_2 B_2(t) + \beta_3 B_3(t) + \beta_4 B_4(t),$$

where (β_l, B_l) are the coefficients and basis functions of the B-spline, respectively $(l = 1, ..., 4)$, noting that $B_l(t)$ is a cubic function of t (Eilers and Marx, 1996; De Boor, 2001). The CAR model is used to capture the spatial random effects η_i. We consider the following general model for the spatial effects η_i,

$$\begin{aligned} &(\eta_1, ..., \eta_I)' \sim N(0, \Sigma_\eta), \\ &\Sigma_\eta = \sigma_\eta^2 (P - \lambda_\eta D)^{-1}, \end{aligned}$$

where P is a $I \times I$ diagonal matrix with elements $P_{ii} = w_{i+}$ where $w_{i+} = \sum_j w_{ij}$ is the number of areas that are adjacent to area i by defining the weights $w_{ij} = 1$ if area i and j are adjacent (shown $i \sim j$) and 0 otherwise; D is a $I \times I$ matrix with elements $D_{ij} = w_{ij}/w_{i+}$ if $i \sim j$ and $D_{ij} = 0$ otherwise (also $D_{ii} = 0$); σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence. Hence, $(P - \lambda_\eta D)$ is then non-singular provided that $|\lambda_\eta| < 1$ (Spiegelhalter *et al.*, 2004). The I full conditional distributions can be also written as

$$\eta_i|\eta_{-i} \sim N\Big(\frac{\lambda_\eta \sum_j w_{ij}\eta_j}{\sum_j w_{ij}}, \frac{\sigma_\eta^2}{\sum_j w_{ij}}\Big), \qquad i = 1, ..., I, \tag{3}$$

where $\eta_{-i} = \{\eta_j : j \neq i, j \sim i\}$.

~~where P is a $I \times I$ diagonal matrix with elements $P_{ii} = 1/h_i$ where h_i is the number of areas that are adjacent to area i; D is a $I \times I$ matrix with elements $D_{ij} = 1/h_i$ if area i and j are adjacent (shown $i \sim j$) and $D_{ij} = 0$ otherwise (also $D_{ii} = 0$); σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence, noting that if $|\lambda_\eta| < 1$, $(I_I - \lambda_\eta D)$ is then non-singular; and I_I is the identity matrix of dimension I (Spiegelhalter *etal.*, 2004).~~

As special case when the conditional spatial dependence λ_η is one, we have the I full conditional distributions as where $\eta_{-i} = \{\eta_j : j \neq i, j \in \partial_i\}$ and $\#\partial_i$ is the number of neighbors of the corresponding area i, and $\bar{\eta}_i$ is the mean of the random effects in the neighborhood of the i th geographic area.

One may define the interaction effect of space and time, δ_{it}, as $\eta_i(t), S_i(t)$, or simply iid Normal distribution, depending on the nature of dataset (Bernardinelli *et al.*, 1992; MacNab and Dean, 2001; Silva *et al.*, 2008; Torabi and Rosychuk, 2012). Note that $\eta_i(t)$ is a CAR model for each specific time t and $S_i(t)$ is a B-spline model for each specific area i. We define $m_0, S_0(t), \eta_{0i}$ and δ_{0it} in a similar manner for the logit part of the model. It depends on the nature of data, in general, whether to use a spatial CAR model for each time or a B-spline for each areal unit. In particular, one can use a spatial CAR model for each time, if the areal trends are different for each time. Similarly, one can use a B-spline model for each areal unit, if the time trends are different for each areal unit. The interaction term should also scientifically sound before incorporating into the model.

3 Likelihood-based estimation

Let $\boldsymbol{y} = (y_{11}, ..., y_{1T}, ..., y_{I1}, ..., y_{IT})'$ be the observed data vector and, conditionally on the random effects, $\boldsymbol{v} = (\eta_1, ..., \eta_I, \delta_{11}, ..., \delta_{IT}, \eta_{01}, ..., \eta_{0I}, \delta_{011}, ..., \delta_{0IT})'$, we assume that the elements of $\boldsymbol{y}$ are independent and drawn from a ZIP distribution (1)-(2) with parameters $\boldsymbol{\alpha}_1 = (m, \beta_1, ..., \beta_4, m_0, \beta_{01}, ..., \beta_{04})$. It is also assumed that distribution for $\boldsymbol{v}$ depends on parameters $\boldsymbol{\alpha}_2$ which include $\lambda_\eta, \sigma^2_\eta, \lambda_{0\eta}, \sigma^2_{0\eta}$ and related parameters from δ_{it} and δ_{0it}. The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$ and prediction of disease incidence over space and time (function of $\boldsymbol{v}$) with corresponding standard errors. The marginal likelihood of data denoted by $L(\boldsymbol{\alpha}; \boldsymbol{y})$ is obtained by integrating conditional probabilities of responses over the distribution of random effects as follows:

$$L(\boldsymbol{\alpha}; \boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{v}, \boldsymbol{\alpha}_1) g(\boldsymbol{v}|\boldsymbol{\alpha}_2) d\boldsymbol{v}, \tag{4}$$

where $f(\cdot)$ is the zero-inflated spatio-temporal model defined as (1)-(2), and $g(\cdot)$ is a multivariate Normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\Sigma_v, \Sigma_v = diag(\Sigma_\eta, \Sigma_\delta, \Sigma_{0\eta}, \Sigma_{0\delta})$, where Σ_δ is the variance-covariance matrix of $(\delta_{11}, ..., \delta_{IT})$ and similarly for $\Sigma_{0\delta}$.

We use the DC method in order to obtain MLE of the parameters of (4). The DC method uses the Bayesian computational approach for frequentist purposes. To understand the idea behind the DC method, imagine a hypothetical situation where the observations $\boldsymbol{y}$ are repeated independently by K different individuals, and all these individuals happen to result in exactly the same set of observations $\boldsymbol{y}$, called $\boldsymbol{y}^{(K)} = (\boldsymbol{y}, \boldsymbol{y}, ..., \boldsymbol{y})$. The likelihood function for the combination of the data from these K independent experiments is then given by $\{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K := L^K(\boldsymbol{\alpha}; \boldsymbol{y})$. Note that this likelihood function has two important features: a) the location of the maximum of this function is exactly equal to the location of the maximum of $L(\boldsymbol{\alpha}; \boldsymbol{y})$ and b) the Fisher information matrix based on this likelihood is K times the Fisher information matrix based on $L(\boldsymbol{\alpha}; \boldsymbol{y})$. Denote $\hat{\boldsymbol{\alpha}}$ as the MLE and $I(\hat{\boldsymbol{\alpha}})$ as corresponding Fisher information matrix based on $L(\boldsymbol{\alpha}; \boldsymbol{y})$. It is assumed that the parameters are identifiable and that there is a unique mode (but possible multiple smaller peaks) to the likelihood function. The posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(K)}$ is then given by

$$\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)}) = \frac{L^K(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(K, \boldsymbol{y})}, \tag{5}$$

where $C(K, \boldsymbol{y}) = \int L^K(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$ is the normalizing constant. The following theorem provides the validity of inference based on the likelihood of K copies of the original data.

Theorem 1: Consider the general models (1) and (2). Under some mild regularity conditions, as K becomes large, the distribution in $\sqrt{K}\Sigma^{-1/2}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}})|\boldsymbol{y}^{(K)}$ converges to a multivariate Normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{I}_p$ which is the identity matrix with the dimension of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$ is the MLE, and Σ is the inverse of the Fisher information matrix for the MLE.

Proof of Theorem 1 is provided in the Appendix.

The Theorem 1 assures that the sample mean vector of the generated random numbers from the posterior distribution (5) provides the MLE of the model parameters $\boldsymbol{\alpha}$, and K times their sample variance-covariance matrix is an estimate of the asymptotic variance-covariance matrix of the MLE $\hat{\boldsymbol{\alpha}}$.

Determining the number of clones K is possible through diagnostics measures and plots (Lele *et al.*, 2010), which are available in the *dclone* package (Sólymos, 2010), a freely available package created for R (R Development Core Team, 2015). We use these criteria to obtain appropriate number of clones in our simulation study and in our application.

3.1 Prediction of random effects

We now need to predict the disease incidence ratio ($DR_{it} = \mu^c_{it}/e_{it}$) at area i and time t, where $\mu^c_{it} = E(y_{it}|\theta_{it}, \mu_{it}, \boldsymbol{\alpha_1})$. Following Hamilton (1986) and Lele *et al.* (2010), based on the MLEs for $\boldsymbol{\alpha}$, the prediction (and prediction interval) of $DR_{it} = dr_{it}$ conditional on the observed data is possible through the following posterior density via Markov chain Monte Carlo (MCMC) sampling:

$$\pi(dr_{it}|\boldsymbol{y}) = \frac{\int f(\boldsymbol{y}|dr_{it}, \boldsymbol{\alpha}_1) g(dr_{it}|\boldsymbol{\alpha}_2) \phi(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, I^{-1}(\hat{\boldsymbol{\alpha}})) d\boldsymbol{\alpha}}{C(\boldsymbol{y})}, \tag{6}$$

where $f(\cdot)$ is the zero-inflated spatio-temporal model defined as (1)-(2), $g(\cdot)$ is a multivariate Normal distribution, $\phi(., \mu, \Sigma)$ denotes a multivariate Normal density with mean μ and variance-covariance Σ, which are equal to the MLE and the inverse of the Fisher information matrix here, and $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha}; \boldsymbol{y}) \pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$ is the normalizing constant.

In this paper, for the DC and hierarchical Bayesian (HB) analysis (i.e. $K = 1$), the independent Normal distribution is assigned for fixed effects with zero mean and variance 10^6, gamma distribution for the inverse of variance components with shape and scale parameter 0.001, and a uniform distribution on $(-1, 1)$ for λ_η. We also use uniform distribution $U(0, 1000)$ instead of the gamma distribution as a prior for the standard deviations (Gelman, 2006). To monitor the convergence of the model parameters, we use several diagnostic methods implemented in the Bayesian output analysis (BOA) program (Smith, 2007) in R. We also use diagnostic methods implemented in the *dclone* package (Sólymos, 2010) to monitor the convergence of the model parameters in terms of number of clones K.

4 Application

4.1 Data analysis

We use a monthly dataset of children (age $\leq$ 20 years) asthma visits to hospital in the Canadian province of Manitoba for during the period April 2006 to March 2010. The population of Manitoba was stable during the study period from 1.18 million people in 2006 to 1.20 million people in 2010, with an average population of children of around 335,000 people. The province consisted of eleven Regional Health Authorities that were responsible for the delivery of health care services. These eleven regions were further sub-divided into 56 Regional Health Authorities Districts (RHADs) and these RHADs are the geographic areas used in our model and all data were linked to these geographic boundaries. For simplicity, we call

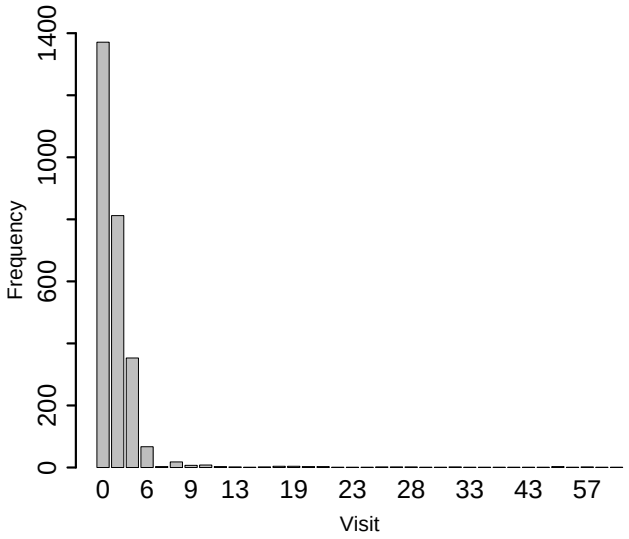

Figure 1. Histogram of number of visits for asthma by children in a month in an area of Manitoba, Canada, from April 2006 to March 2010

these areas 1,2,...,56. The number of children with asthma visits totaled 4,948 over the study period with mean and median number of monthly asthma visits per area of 2 and 0 (range 0 to 69), respectively. The yearly areal children population sizes varied from 290 to 173,400 people, with mean and median numbers of 5,974 and 2,382 people, respectively. The largest population was in area 56, while area 42 had the least population. Figure 1 gives a histogram of number of visits for asthma by children in a month in an area of Manitoba for all 2,688 ($= I \times T = 56 \times 48$) count values with more than 50% zeros which clearly shows the need of a zero-inflation model. We also present the provincial rate of children asthma visits over time. Figure 2 shows the overall crude rates, $\sum_{i=1}^{I} y_{it} / \sum_{i=1}^{I} n_{it}$, of children asthma visits to hospital over time where n_{it} is the population at risk at area i and month t.

The following model is found useful in our exploration of the data:

$$\log(\mu_{it}) = \log(e_{it}) + \eta_i + S(t) + \delta_{it},$$

$$logit(\theta_{it}) = \eta_{0i} + S_0(t), \tag{7}$$

where $S(t)$ is a cubic B-spline, η_i is a CAR model, $\delta_{it} \sim N(0, \sigma_\delta^2), \eta_{0i} \sim N(0, \sigma_{\eta 0}^2)$, and $S_0(t) \sim N(0, \sigma_t^2)$. Note that the expected number of asthma visits (e_{it}) is adjusted with respect to gender by

$$e_{it} = \sum_{l=1}^{2} n_{itl} \frac{y_l}{n_l},$$

where n_{itl} is the population at risk at area i and time t and gender l, $y_l = \sum_{i=1}^{I} \sum_{t=1}^{T} y_{itl}$ where y_{itl} is the number of children asthma visits at area i and time t and gender l, and similarly $n_l = \sum_{i=1}^{I} \sum_{t=1}^{T} n_{itl}$. We first fit the model (7) to the dataset of children asthma visits using the DC and HB methods. Following Ponciano *et al.* (2009) and Torabi (2014), we use model selection based on the information criteria for this purpose. In particular, to compare two models, one can write $AIC_1 - AIC_2 = -2ln(\frac{\hat{L}_1}{\hat{L}_2}) + 2(d_1 - d_2)$,

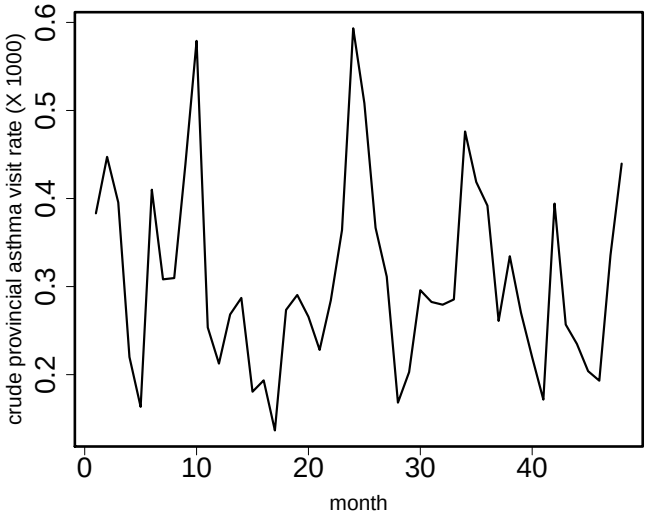

Figure 2. Provincial children asthma visits rate over time in the province of Manitoba, Canada, during the period April 2006 to March 2010

where AIC stands for Akaike Information Criteria (e.g., $AIC_1 = 2d_1 - 2ln(L_1)$), $\hat{L}_i$ is the complete likelihood function evaluated at MLE $\hat{\alpha}$ for model $i(i = 1, 2)$, d_1 and d_2 are the number of estimated parameters under models 1 and 2, respectively (Burnham and Anderson, 2002; Ponciano *et al.*, 2009). We define the spatial model (only η_i and η_{0i} in (7)) as model (i), the temporal model (only $S(t)$ and $S_0(t)$) in (7) as model (ii) , and the model (7) as model (iii). We then found that $AIC_i - AIC_{iii} = 4.32$, $AIC_{ii} - AIC_{iii} = 6.12$. The AIC differences greater than two are generally thought to be significant, and differences greater than three are very signifiant (Burnham and Anderson, 2002; Taper, 2004; Ponciano *et al.*, 2009). We conclude that the spatio-temporal model (7) provides a better description of the data than models with only spatial or only temporal effects. Table 1 reports the model parameters estimates and corresponding standard errors for the both DC and HB approaches. It seems that the standard errors in the DC approach are smaller than the corresponding values in the HB method. We also observe that the HB method performs differently for the gamma and uniform prior distributions unlike the DC method. For this application, the number of clones was $K = 5$ to obtain MLE, and the number of iterations for convergence of the model parameters in DC and HB methods was about 40,000. In particular, to check the convergence of the DC approach, we need to calculate the largest eigenvalue of the posterior variance-covariance matrix, or to calculate mean squared error and another correlation-like fit statistic based on a Chi-squared approximation. The maximum eigenvalue reflects the degenerateness of the posterior distribution, while the two fit measures reflect if the Normal approximation is adequate. All three statistics should converge to zero as the number of clones increases. If this happens, different prior specifications are no longer influencing the results (Lele *et al.*, 2007, 2010). These are conveniently collected by the *dcdiag* function in *dclone* package in R. In the case of $K = 5$, the maximum eigenvalue, mean squared error, and correlation-like fit statistic are 0.02, 0.002, and 0.001, respectively. To also further investigate the behavior of the convergence of the DC approach, we provide a plot of posterior variances for different values of K (Figure 3).

We also study the sensitivity of our results to the number of knots. To that end, we consider different number of knots (3, 6, 10) for our $S(t)$ defined in (7) for the cases of MLE and HB (using uniform distribution as prior for the standard deviations), noting that we used 3 knots as default for the all results in

Table 1. Parameter estimates (and standard errors), zero-inflated spatio-temporal model for MLE and HB methods, children asthma visits to hospital in the province of Manitoba, Canada, during the period April 2006 to March 2010

	Estimate (Standard error)			
Parameter	MLE Gamma distribution	MLE Uniform distribution	HB Gamma distribution	HB Uniform distribution
β_1	-0.110 (0.167)	-0.112 (0.149)	-0.100 (0.215)	-0.108 (0.173)
β_2	-0.326 (0.206)	-0.325 (0.219)	-0.526 (0.272)	-0.334 (0.212)
β_3	-0.010 (0.210)	-0.011 (0.198)	0.185 (0.278)	-0.011 (0.223)
β_4	-0.282 (0.155)	-0.283 (0.145)	-0.418 (0.206)	-0.285 (0.168)
σ^2_η	0.816 (0.190)	0.814 (0.178)	0.905 (0.250)	0.823 (0.200)
λ_η	0.975 (0.029)	0.973 (0.027)	0.973 (0.029)	0.965 (0.032)
σ^2_δ	0.117 (0.013)	0.116 (0.011)	0.313 (0.031)	0.121 (0.021)
$\sigma^2_{\eta 0}$	0.926 (0.218)	0.925 (0.199)	3.118 (0.643)	0.923 (0.232)
σ^2_t	0.213 (0.003)	0.214 (0.003)	0.003 (0.003)	0.224 (0.004)

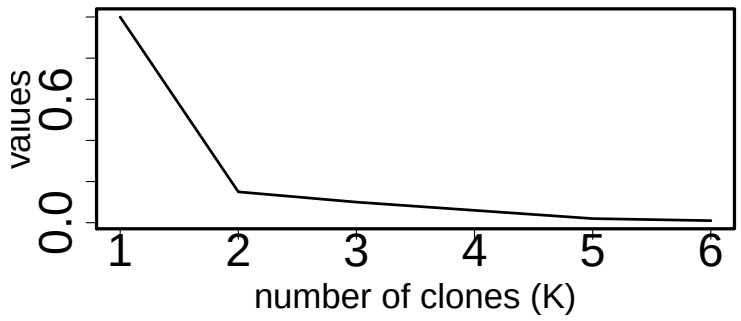

Figure 3. Data cloning convergence diagnostics for the children asthma visits to hospital in the province of Manitoba, Canada. The standardized eigenvalues converge to zero as the number of clones (K) increases

this manuscript unless otherwise stated. As shown in Table 2, the results are not sensitive to the number of knots in our data analysis.

For a diagnostic analysis, we adopt and calculate the *deviance residual* (McCullagh and Nelder, 1989) as

$$d_{it} = sgn(y_{it} - \hat{\mu}^c_{it})\Big[2\big\{y_{it}\log(\frac{y_{it}}{\hat{\mu}^c_{it}}) - y_{it} + \hat{\mu}^c_{it}\big\}\Big]^{1/2},$$

where

$$sgn(z) = \begin{cases} 1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases},$$

note that the deviance residual formula was originally proposed for the generalized linear models. Figure 4 gives the residuals versus predicted diagnostic plot based on the MLE approach. It is clear from Figure 4

Table 2. Sensitivity of our results to the number of knots (3, 6, and 10) for the cases of MLE and HB (using uniform distribution as prior for the standard deviations), children asthma visits to hospital in the province of Manitoba, Canada, during the period April 2006 to March 2010

Parameter	Estimate (Standard error)					
	MLE			HB		
	3 knots	6 knots	10 knots	3 knots	6 knots	10 knots
β_1	-0.112 (0.149)	-0.113 (0.165)	-0.111 (0.169)	-0.108 (0.173)	-0.125 (0.175)	-0.118 (0.174)
β_2	-0.325 (0.219)	-0.327 (0.204)	-0.326 (0.209)	-0.334 (0.212)	-0.376 (0.223)	-0.344 (0.210)
β_3	-0.011 (0.198)	-0.010 (0.211)	-0.011 (0.208)	-0.011(0.223)	-0.029 (0.243)	-0.021 (0.229)
β_4	-0.283 (0.145)	-0.281 (0.156)	-0.282 (0.153)	-0.285 (0.168)	-0.349 (0.183)	-0.265 (0.172)
σ^2_η	0.814 (0.178)	0.815 (0.191)	0.816 (0.183)	0.823 (0.200)	0.845 (0.232)	0.803 (0.226)
λ_η	0.973 (0.027)	0.974 (0.029)	0.973 (0.027)	0.965 (0.032)	0.973 (0.034)	0.943 (0.032)
σ^2_δ	0.116 (0.011)	0.114 (0.014)	0.115 (0.013)	0.121 (0.021)	0.145 (0.031)	0.109 (0.021)
$\sigma^2_{\eta 0}$	0.925 (0.199)	0.923 (0.217)	0.924 (0.215)	0.923 (0.232)	0.943 (0.243)	0.915 (0.235)
σ^2_t	0.214 (0.003)	0.212 (0.003)	0.213 (0.003)	0.224 (0.004)	0.256 (0.003)	0.238 (0.004)

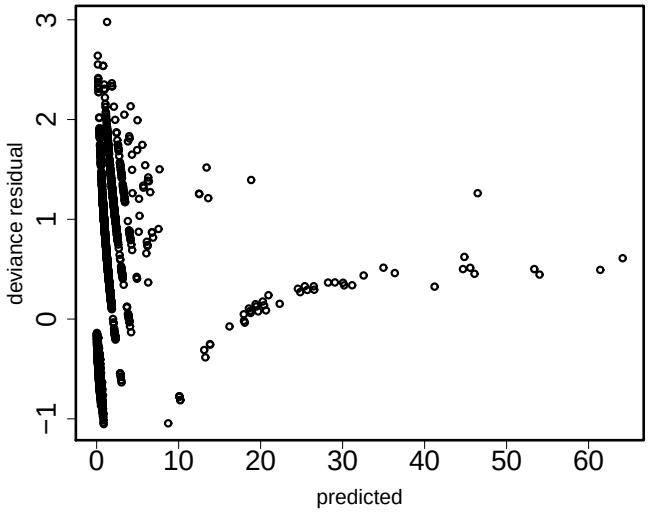

Figure 4. The deviance residuals versus predicted diagnostic plot of children asthma visits based on MLE approach

that there is no serious lack of fit in our model; noting that those observations with large predicted values in Figure 4 belong to Winnipeg health region (area 56, largest population) with relatively large number of children asthma visits to hospital, and there is a stripy effect in the left side which could be due to the excess of zeros or small counts. Note that very small and positive values were given to y_{it} if they were 0 to avoid the error in calculation of d_{it}.

One of the main features of DC is also the ability to predict the random effects. To have better understanding of the estimated spatial risk profile, we obtain the adjusted children asthma visits ratio, μ^c_{it}/e_{it}, using DC, which provides a spatial risk profile. Figure 5 presents maps of the estimated spatial effects based on the fitted model, where the areal risk factor of children asthma visits corresponds to, for example,

the months of April in 2006 and 2008. The overall spatial pattern suggests that some areas in the south and many areas in the north-central part of the province have relatively high children asthma visits ratio. Generally, the spatial pattern does not change much over time; although some RHADs had higher children asthma visits ratio estimate in 2008 compared to 2006 (e.g., RHADs 10, 32, and 40) and some RHADs had lower children asthma visits ratio estimate in 2008 compared to 2006 (e.g., RHADs 29, 30, and 38). More investigation may be needed to explore the reasons for seemingly higher children asthma visits in these areas compared to other parts of the province.

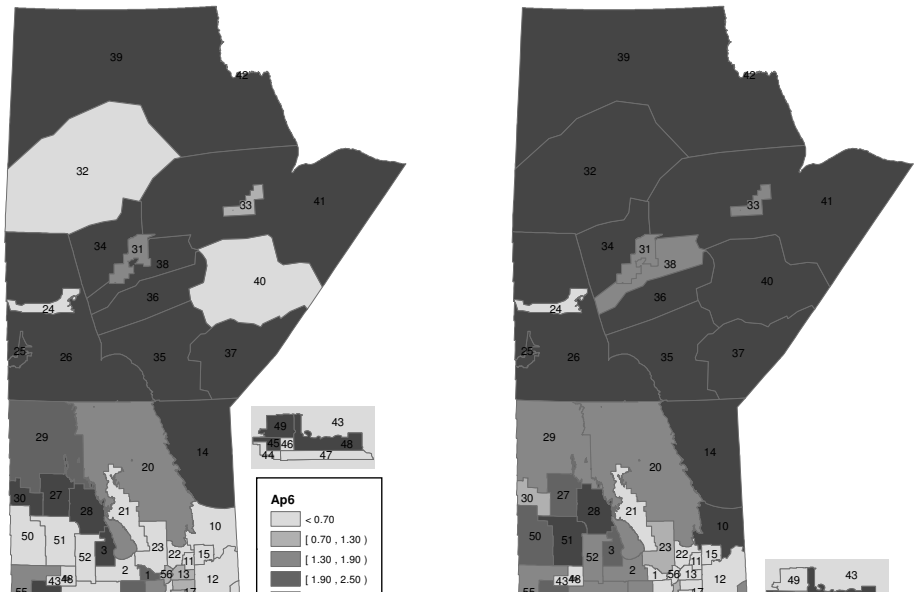

Figure 5. Adjusted children asthma visits ratio for the spatial effects of the areal asthma risks in April for selected years 2006 and 2008; Manitoba children asthma data, April 2006-March 2010

We also provide the areal children asthma visits ratio estimate obtained from fitting the spatio-temporal mixed model given by $\exp(S(t) + \eta_i + \delta_{it})$. Figure 6 plots the fitted children asthma visits ratio with corresponding 95% prediction intervals, for example, for areas (regions) 2, 5, 42, and 56 using the DC method. The crude ratio estimates are y_{it}/e_{it}, and are also plotted in Figure 6. We indeed chose two regions with extreme population sizes; region 42 with least population and region 56 with largest population. As expected in Figure 6, our children asthma visits ratio estimates provide smoothed estimates while crude ratios are very unstable over time particularly for region 42 with low population size. In general, a specific pattern in estimated log ratio over time for an area would suggest that the underlying children asthma visits rate in that area has also the same pattern relative to the provincial average.

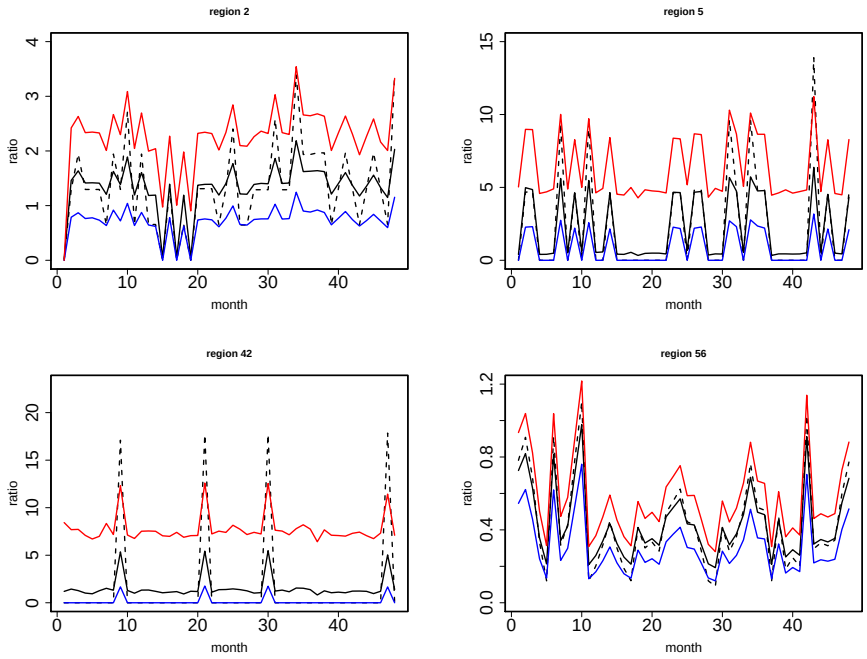

Figure 6. Fitted children asthma visits ratio to hospital for selected areas (regions) 2, 5, 42, and 56 during the period April 2006 to March 2010. The solid black line represents fitted ratios with blue and red lines as 95% prediction bands; the dashed line is crude ratio

4.2 Simulation study

We conduct a simulation study to evaluate and compare the performance of ML estimates, via DC approach, with the HB method using a scenario similar to our children asthma dataset. More specifically, data are generated from the model (7) with the parameters set close to those obtained in the analysis of the children asthma dataset; $\beta_1, \beta_2, \beta_3, \beta_4, \sigma_\eta^2, \lambda_\eta, \sigma_\delta^2, \sigma_{\eta 0}^2$ and σ_t^2 are listed in Table 3. The neighborhood structure and the population sizes are exactly as for the asthma dataset, noting that we use n_{it} (population at area i and time t) rather than e_{it} in the model (7). Estimates are obtained using DC and HB analyses of $1,000$ datasets generated from the model (7) where the non-informative priors are used for the fixed effects, and gamma distribution for the variance components.

Table 3 presents the bias values of the model parameters estimate, the empirical standard errors of model parameters estimate, and the model-based standard errors of the estimate parameters. In the case of DC method, the estimates are fairly unbiased, and it seems that their standard errors are estimated reasonably well unlike the HB approach. Overall, it seems that DC approach, which yields to MLE, provides better point estimates and standard errors for this data analysis compared to the HB approach.

We also study the performance of the estimated disease rate, $DR_{it} = \exp(S(t) + \eta_i + \delta_{it})$, by providing corresponding coverage probabilities using the DC and HB approaches. Table 4 shows the average coverage probabilities of prediction intervals of the disease rates (DR_{it}) over areas and time for the both DC and HB methods. The proposed approach also performs very well (and slightly better than the HB method) in terms of average coverage probabilities of the disease rates.

Table 3. Bias values, model-based standard errors, and empirical standard errors of the estimate parameters; MLE and HB methods based on 1,000 simulated datasets

Parameter	MLE			HB		
	Bias	Standard error		Bias	Standard error	
		MLE	Empirical		HB	Empirical
$\beta_1 = -0.10$	-0.0003	0.165	0.169	-0.0011	0.213	0.171
$\beta_2 = -0.30$	0.0004	0.197	0.204	0.0014	0.261	0.213
$\beta_3 = -0.01$	-0.0004	0.200	0.215	0.0006	0.268	0.230
$\beta_4 = -0.30$	-0.0002	0.158	0.153	-0.0011	0.209	0.155
$\sigma^2_\eta = 0.80$	0.0003	0.187	0.186	0.0013	0.246	0.191
$\lambda_\eta = 0.95$	0.0001	0.027	0.026	0.0012	0.026	0.031
$\sigma^2_\delta = 0.10$	0.0003	0.012	0.011	0.0014	0.030	0.021
$\sigma^2_{\eta 0} = 0.90$	0.0002	0.216	0.215	0.0008	0.641	0.222
$\sigma^2_t = 0.20$	0.0003	0.003	0.003	0.0007	0.003	0.005

Table 4. Coverage probabilities of confidence intervals of the DR_{it} averaged over areas and time ($\overline{DR}$), with confidence coefficients $0.90, 0.95, 0.98$, and 0.99 for the MLE and HB methods based on 1,000 simulated datasets

Parameter	DC				HB			
	Confidence coefficient				Confidence coefficient			
	0.90	0.95	0.98	0.99	0.90	0.95	0.98	0.99
$\overline{DR}$	0.894	0.946	0.978	0.989	0.892	0.943	0.975	0.986

5 Conclusion

In many applications, there are situations that spatial and temporal count data have excess zeros. For instance, in our application, more than 50% of our monthly children asthma visits to hospital per area had 0 values. In this paper, we proposed a zero-inflated spatio-temporal model to account for spatial and temporal variations from excess zero counts. The model accommodated a CAR model for the spatial random effects and B-spline smoothing over the temporal effects. We also proposed to use the DC method which yields to MLE to estimate the model parameters, and also to predict the disease incidence ratios over space and time. The advantage of DC approach is that the prediction (and prediction intervals) of the smoothing incidence ratios over space and time is easily obtained.

We adjusted our expected number of children asthma visits to hospital by an important factor of gender. The model can be also easily extended to include some covariates directly, which may be required for some applications. Overall, it was suggested by the model estimates that the high children asthma incidences ratio were mainly located in some parts in the south and many parts in the north-central part of the province. These findings may represent real increases or different distributions of important covariates that are unmeasured and unadjusted for in our modeling. Further investigation is needed to explore these findings.

Acknowledgements

I would like to thank the Editor, an Associate Editor, and a referee for constructive comments and suggestions, which led to the improvement of the paper. This work was supported by a grant from the Natural

Sciences and Engineering Research Council of Canada (NSERC).

Conflict of Interest
The author has declared no conflict of interest.

Appendix

Proof of Theorem 1: Proof of this Theorem closely follows the lines of Lele *et al.* (2010). Let $L(\boldsymbol{\alpha};\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{v},\boldsymbol{\alpha}_1)g(\boldsymbol{v}|\boldsymbol{\alpha}_2)d\boldsymbol{v}$ is the likelihood function of the data vector $\boldsymbol{Y} = (\boldsymbol{Y}_{11},...,\boldsymbol{Y}_{IT})^T$. We also assume that there is a bounded function as a function of $\boldsymbol{\alpha}$. Let $\pi(\boldsymbol{\alpha})$ is the prior distribution on the parameter space $\boldsymbol{\vartheta}$. Let $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L^K(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(K,\boldsymbol{y})}$ where $C(K,\boldsymbol{y}) = \int L^K(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha} < \infty$. Following Lele *et al.* (2010), we assume the following three assumptions: (A1) the function $L(\cdot)$ has a local maximization at $\boldsymbol{\alpha}_\infty$ and $L(\boldsymbol{\alpha}_\infty) > 0$ and $\pi(\boldsymbol{\alpha}_\infty) > 0$ where $\boldsymbol{\alpha}_\infty(= \hat{\boldsymbol{\alpha}})$ is defined as the MLE. (A2) the function $\pi(\cdot)$ is continuous at $\boldsymbol{\alpha}_\infty$ and function $L(\cdot)$ has continuous second derivatives in a neighborhood of $\boldsymbol{\alpha}_\infty$ and $D^2L(\boldsymbol{\alpha}_\infty) = \frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T} log L(\boldsymbol{\alpha};\boldsymbol{y})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_\infty}$ is strictly negative definite. (A3) for any $\delta > 0$, we have $\gamma(\delta) := sup\{L(\boldsymbol{\alpha};\boldsymbol{y}) : ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty|| > \delta\} < L(\boldsymbol{\alpha}_\infty)$. We also define $\Sigma^{1/2} = \{-D^2L(\boldsymbol{\alpha}_\infty)\}^{-1/2}$ and for $\delta > 0$ there is $N(\delta) := \{\boldsymbol{\alpha} : ||\Sigma^{-1/2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty)|| < \delta\}$. Since $\Sigma^{1/2}$ is positive definite, this defines a system of neighborhoods of $\boldsymbol{\alpha}_\infty$. Also, let $\boldsymbol{\alpha}_K$ is a random variable on R^p with density function $\pi_K(\cdot)$ and define standardized variable $\Psi_K = \sqrt{K}\Sigma^{-1/2}(\boldsymbol{\alpha}_K - \boldsymbol{\alpha}_\infty)$ with density function $h_K(\boldsymbol{\alpha}) = \frac{|\Sigma^{1/2}|}{K^{p/2}}\pi_K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})$ where p is the dimension of $\boldsymbol{\alpha}$. Without loss of generality, we assume that $L(\boldsymbol{\alpha}_\infty) = 1$ which is a standardized likelihood function where the computation of the posterior distribution $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y})$ will be invariant to such standardization as $L^K(\boldsymbol{\alpha};\boldsymbol{y})$ is involved in both numerator and denominator of $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y})$. Hence, $\Sigma^{1/2} = \{-D^2L(\boldsymbol{\alpha}_\infty)\}^{-1/2}$ is the square root of the inverse of the Fisher information matrix due to $D^2L(\boldsymbol{\alpha}_\infty) = D^2 \log L(\boldsymbol{\alpha}_\infty)$.

The proof of Theorem 1 is based on two main results: we first show that under the assumptions (A1) and (A2), $L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})$ converges to $\exp(-||\boldsymbol{\alpha}||^2/2)$ uniformly on bounded sets of $\boldsymbol{\alpha}$ as K goes to ∞. To that end, we fix $\delta_0 > 0$ so small such that $D^2L(\boldsymbol{\alpha})$ is continuous on the neighborhood $N(\delta_0)$. For every $\boldsymbol{\alpha}$ in this neighborhood, there is some $\boldsymbol{\alpha}^+$ on the line segment joining $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_\infty$ so that by using Taylor's expansion, we can have

$$\begin{aligned} L(\boldsymbol{\alpha}) &= L(\boldsymbol{\alpha}_\infty) + DL(\boldsymbol{\alpha}_\infty)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty) + \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty)^T(D^2L(\boldsymbol{\alpha}^+))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty) \\ &= 1 - \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty)^T(-D^2L(\boldsymbol{\alpha}^+))(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\infty). \end{aligned} \tag{8}$$

So, for large K, we have $\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha}$ in $N(\delta_0)$ and also $L(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha}) = 1 - \frac{\boldsymbol{\alpha}^T(\Sigma^{1/2})^T\{-D^2L(\boldsymbol{\alpha}_K)\}\Sigma^{1/2}\boldsymbol{\alpha}}{2K}$ for some $\boldsymbol{\alpha}_K$ on the line segment joining $\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_\infty$. We can also choose $\delta(\epsilon) < \delta_0$ (for $\epsilon > 0$) so small for $\boldsymbol{\alpha} \in N(\delta(\epsilon))$ such that $D^2L(\boldsymbol{\alpha})$ is negative definite and $||(\Sigma^{1/2})^T(-D^2L(\boldsymbol{\alpha}))\Sigma^{1/2} - \boldsymbol{I}_p|| \leq \epsilon$. Also, in general, for $0 \leq x, y \leq K$, we have: $|(1 - \frac{x}{K})^K - (1 - \frac{y}{K})^K| \leq |x - y|$ and $|(1 - \frac{y}{K})^K - \exp(-y)| \leq \frac{y^2}{K}$. For $M > 1$ and $0 < \epsilon < 1$ and let $K \geq \max((\frac{M}{\delta(\epsilon)})^2, M^2)$. Hence, for $||\boldsymbol{\alpha}|| < M$ we have $\boldsymbol{\alpha}_K \in N(\delta(\epsilon))$, so with $x = \frac{1}{2}\boldsymbol{\alpha}^T(\Sigma^{1/2})^T(-D^2L(\boldsymbol{\alpha}_K))\Sigma^{1/2}\boldsymbol{\alpha}$ and $y = ||\boldsymbol{\alpha}||^2/2$, we get

$$|L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha}) - \exp(-\frac{||\boldsymbol{\alpha}||^2}{2})| \leq \frac{\epsilon M^2}{2} + \frac{M^4}{4K}, \tag{9}$$

which shows that $L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})$ converges to $\exp(-||\boldsymbol{\alpha}||^2/2)$ uniformly on bounded sets of $\boldsymbol{\alpha}$ as K goes to ∞. There are some immediate results based on (9). We can now conclude that $\pi(\boldsymbol{\alpha}_\infty +$

$\frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})$ converges to $\pi(\boldsymbol{\alpha}_\infty)\exp(-||\boldsymbol{\alpha}||^2/2)$ uniformly on bounded sets. We can also have $\pi(\boldsymbol{\alpha}_\infty)|\Sigma^{1/2}|(2\pi)^{p/2} \leq \liminf_K C(K,\boldsymbol{y})K^{p/2}$ where there is a constant $D > 0$ such that $\frac{1}{C(K,\boldsymbol{y})} \leq DK^{p/2}$.

The second main result is to show the following three are equivalent: (a) $\Psi_K \Rightarrow N(\mathbf{0}, \boldsymbol{I}_p)$ which is convergence in distribution to a Normal random variable. (b) the density $h_K(\cdot)$ converges point-wise to a multivariate standard Normal density function. In other words, $C(K,\boldsymbol{y})K^{p/2} \longrightarrow \pi(\boldsymbol{\alpha}_\infty)|\Sigma^{1/2}|(2\pi)^{p/2}$. (c) $\boldsymbol{\alpha}_K \Rightarrow \delta_{\boldsymbol{\alpha}_\infty}$ where $\delta_{\boldsymbol{\alpha}_\infty}$ indicates a degenerate distribution at $\boldsymbol{\alpha}_\infty$. To show $(a) \Rightarrow (b)$, we can write the density $h_K(\cdot)$ as:

$$h_K(\boldsymbol{\alpha}) = \frac{|\Sigma^{1/2}|}{K^{p/2}C(K,\boldsymbol{y})}\pi(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha}).$$

By (a), we have the following on the bounded Borel set B :

$$\frac{1}{(2\pi)^{p/2}}\int_B \exp(-||\boldsymbol{\alpha}||^2/2)d\boldsymbol{\alpha} = \lim_K \frac{|\Sigma^{1/2}|}{C(K,\boldsymbol{y})K^{p/2}}\int_B \pi(\boldsymbol{\alpha}_\infty+\frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})L^K(\boldsymbol{\alpha}_\infty+\frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})d\boldsymbol{\alpha}.$$

On the other hand, using (9):

$$\lim_K \int_B \pi(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})L^K(\boldsymbol{\alpha}_\infty + \frac{1}{\sqrt{K}}\Sigma^{1/2}\boldsymbol{\alpha})d\boldsymbol{\alpha} = \pi(\boldsymbol{\alpha}_\infty)\int_B \exp(-||\boldsymbol{\alpha}||^2/2)d\boldsymbol{\alpha}.$$

Hence, we can conclude that $C(K,\boldsymbol{y})K^{p/2} \longrightarrow \pi(\boldsymbol{\alpha}_\infty)|\Sigma^{1/2}|(2\pi)^{p/2}$ as K goes to ∞, and consequently $h_K(\boldsymbol{\alpha}) \longrightarrow \frac{1}{(2\pi)^{p/2}}\exp(-||\boldsymbol{\alpha}||^2/2)$. We also now show $(c) \Rightarrow (b)$. For any $\epsilon > 0$, we can find $\delta > 0$ such that $\boldsymbol{\alpha} \in N(\delta)$ which implies (using (8)): $L(\boldsymbol{\alpha}) < 1 - \frac{1}{2}(1-\epsilon)(\boldsymbol{\alpha}-\boldsymbol{\alpha}_\infty)^T\Sigma^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_\infty)$ and $\pi(\boldsymbol{\alpha}) \leq (1+\epsilon)\pi(\boldsymbol{\alpha}_\infty)$. Also, from (c), we can assume that K is large enough to have $1 - \epsilon \leq \int_{N(\delta)} \pi_K(\boldsymbol{\alpha})d\boldsymbol{\alpha}$. We can then show that $C(K,\boldsymbol{y})K^{p/2} \leq (1-\epsilon)^{-1}(\pi(\boldsymbol{\alpha}_\infty)+\epsilon)|\Sigma^{1/2}|(2\pi)^{p/2}$. So, by letting $K \longrightarrow \infty$ and $\epsilon \longrightarrow 0$, we have $\limsup_K C(K,\boldsymbol{y})K^{p/2} \leq \pi(\boldsymbol{\alpha}_\infty)|\Sigma^{1/2}|(2\pi)^{p/2}$. Other parts (e.g., $(a) \Rightarrow (c)$ or $(b) \Rightarrow (a)$) are either obvious or easy to show. Then, an immediate result is that $\boldsymbol{\alpha}_K \Rightarrow \delta_{\boldsymbol{\alpha}_\infty}$ which is also based on the three assumptions (A1-A3).

Hence, we can get the main result of the convergence of data-cloning algorithm $\Psi_K = \sqrt{K}\Sigma^{-1/2}(\boldsymbol{\alpha}_K - \boldsymbol{\alpha}_\infty) \Rightarrow N(\mathbf{0}, \boldsymbol{I}_p)$ under the three assumptions (A1)-(A3), which completes the proof of Theorem 1.

References

[1] Agarwal, D.K. Gelfand, A.E. and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* **9**, 341–355.

[2] Böhning, D. (1998). Zero-Inflated Poisson models and C.A.MAN: A tutorial collection of evidence. *Biometrical Journal* **40**, 833–843.

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

[4] Burnham, K.P. and Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. 2nd edn, Springer-Verlag, New York.

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

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

[7] De Boor, C. (2001). *A Practical Guide to Splines*, Springer-Verlag, New York.

[8] Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Sciences* **11**, 89–121.

[9] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

[10] Gilks, W.R. Richardson, S. and Spiegelhalter, D.J. (eds.). (1996). *Markov Chain Monte Carlo in Practice*, Springer - Verlag, New York.

[11] Hamilton, J.D. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics* **33**, 387–397.

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

[13] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

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

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

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

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

[18] Martinez-Beneito, M.A. Lopez-Quilez, A. and Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* **27**, 2874–2889.

[19] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

[20] Nieto-Barajas, L.E. and Bandyopadhyay, D. (2013). A zero-inflated spatial gamma process model with applications to disease mapping. *Journal of Agricultural, Biological and Environmental Statistics* **18**, 137–158.

[21] Ponciano, J.M. Taper, M.L. Dennis, B. and Lele, S.R. (2009). Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* **90**, 356–362.

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

[23] R Development Core Team. (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org.

[24] Rathbun, S.L. and Fei, S. (2006). A spatial zero-inflated Poisson regression model for oak regeneration. *Environmental and Ecological Statistics* **13**, 409–426.

[25] Silva, G.L. Dean, C.B. Niyonsenga, T. and Vanasse, A. (2008). Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine* **27**, 2381–2401.

[26] Smith, B.J. (2007). *BOA User Manual (version 1.1.7)*, Department of Biostatistics, College of Public Health, University of Iowa, Ames.

[27] Sólymos, P. (2010). dclone: data cloning in R. *The R Journal* 2, 29–37.

[28] Spiegelhalter, D. Thomas, A. Best, N. and Lunn, D. (2004). *WinBUGS version 1.4 User Manual.* MRC Biostatistics unit, Institute of Public Health, London.

[29] Taper, M.L. (2004). Model identification from many candidates. In The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations, M.L.Taper and S.R. Lele (eds), The University of Chicago Press, Chicago, 448–524.

[30] Torabi, M. (2012). Hierarchical Bayes estimation of spatial statistics for rates. *Journal of Statistical Planning and Inference* **142**, 358–365.

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

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

[33] Torabi, M. and Rosychuk, R.J. (2011). Spatio-temporal modeling using spline for disease mapping: Analysis of childhood cancer trends. *Journal of Applied Statistics* **38**, 1769–1781.

[34] Torabi, M. and Rosychuk, R.J. (2012). Hierarchical Bayesian spatiotemporal analysis of childhood cancer trends. *Journal of Geographical Analysis* **44**, 109–120.

[35] Tse, S.K., Chow, S.C., Lu, Q. and Cosmatos, D. (2009). Testing homogeneity of two zero-inflated Poisson populations. *Biometrical Journal* **51**, 159–170.

[36] Ver Hoef, J.M. and Jansen, J.K. (2007). Space-time zero-inflated count models of Harbor seals. *Environmetrics* **18**, 697–712.

[37] Ugarte, M.D., Ibanez, B. and Militino, A.F. (2004). Testing for Poisson zero inflation in disease mapping. *Biometrical Journal* **46**, 526–539.

[38] Walker, A.M. (1969). On the asymptotic behavior of posterior distributions. *Journal of Royal Statistical Society, Series B* **31**, 80–88.

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

[40] Wan, W-Y. and Chan, J.S.K. (2009). A new approach for handling longitudinal count data with zero-inflation and overdispersion: Poisson geometric process model. *Biometrical Journal* **51**, 556–570.

[41] Welsh, A.H. Cunningham, R.B. Donnelly, C.F. and Lindenmayer, D.B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modeling* **88**, 297–308.

[42] Wikle, C.K. and Anderson, C.J. (2003). Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. *Journal of Geophysical Research* **108**, 9005, doi:10.1029/2002JD002806.

[43] Yau, K.K.W., Wang, K. and Lee, A.H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal* **45**, 437–452.

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

