Spatio-temporal modeling for disease mapping using CAR and B-spline smoothing

Mahmoud Torabi *†

In this paper, generalized additive mixed models are constructed for the analysis of geographical and temporal variability of disease ratios. In this class of models, spatio-temporal models that use conditionally autoregressive smoothing across the spatial dimension and B-spline smoothing over the temporal dimension are considered. The frequentist analysis of these complex models is computationally difficult. 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 and equally computationally convenient. We propose to use data cloning, which yields to maximum likelihood estimation, to conduct frequentist analysis of spatio-temporal modeling of disease ratios. The advantages of data cloning approach are that the non-estimable parameters are flagged automatically and prediction (and prediction intervals) of the smoothing incidence ratios over space and time are easily obtained. We illustrate this approach using a real dataset of yearly childhood asthma visits to hospital in the province of Manitoba, Canada, during 2000-

*Correspondence to: Mahmoud Torabi, Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, R3E 0W3, Canada. E-mail: torabi@cc.umanitoba.ca

†Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

2010. The performance of the data cloning approach is also studied through a simulation study.

Keywords: Bayesian computation; conditional autoregressive; geographic epidemiology; hierarchical models; random effects; spline

1. INTRODUCTION

Mapping rates (or ratios) is essentially a way of describing the spatial and sometimes spatio-temporal distribution of rates over a region. Such distributions display the geographic variation in mortality or disease incidence and are very important for epidemiological and health-policy purposes. The idea behind developments on spatial and spatio-temporal modeling of disease ratios is essentially to model variations in true ratios and better separate systematic variability from random noise, a component that usually overshadows crude ratio maps. Maps of regional morbidity and mortality ratios over time are useful tools in determining spatial and temporal patterns of disease and also for targeting resources. Disease incidence and mortality ratios may differ substantially across geographical regions. A reliable estimate of the underlying disease risk is usually provided by *borrowing strength* from neighboring geographic sub-areas.

Poisson regression is commonly used for the analysis of disease cases, which implicitly assumes that the cases 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 and 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.

The temporal smoothing with random effects 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 existing hierarchical Bayesian spatial models to account for temporal random effects and spatio-temporal interactions. A unified approach for a Bayesian analysis of incidence or mortality data 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 effects and B-spline smoothing over the temporal effects. Martinez-Beneito *et al.* (2008) suggested an AR spatio-temporal model based on Bayesian time series and spatial modeling to link information in time and space. In some contexts, the underlying rates may change over seasons within a given year. Torabi and Rosychuk (2010) proposed spatio-temporal models that use conditional AR (CAR) smoothing across the spatial effects, AR smoothing over the temporal effects, and also use a smoothing function to account for seasonal effects. Torabi (2012a) proposed spatio-temporal models that use AR smoothing across the spatial effects, random walk smoothing over the temporal effects, and a smoothing function to account for seasonal effects.

There are many different ways to perform inference in mixed models, however, the frequentist approach has been computationally difficult particularly for our generalized additive mixed model (GAMM); see Section 2 for more details. Consequently, many approximate approaches have been proposed in last two decades such as generalized estimating equations (Liang and Zeger, 1986; Prentice and Zhao, 1991; Torabi and Rosychuk, 2010) and penalized quasi-likelihood (PQL) (Breslow and Clayton, 1993; Torabi and Rosychuk, 2011) among other approaches. However, the maximum likelihood estimation (MLE) approach has been ignored due to high dimension of spatial and temporal

random effects.

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. For example, non-informative priors are often improper, and the use of an improper prior may lead to an improper posterior; moreover, the conditions for posterior property are not easily determined, particularly in spatial models with random effects. The inferences may also depend on the choice of prior (Efron, 1986; Johnson, 1991).

Recently, Lele *et al.* (2007) introduced an alternative frequentist approach, called data cloning (DC), to compute the MLE and their standard errors for general hierarchical models. Similar to the Bayesian approach, DC avoids high dimensional numerical integration and requires neither maximization nor differentiation of a function. Extending this work to the generalized linear mixed model (GLMM) situation, Lele *et al.* (2010) described an approach to compute prediction and prediction intervals for the random effects. Torabi (2012b) also extended the DC approach to the GLMM with two components of dispersion. Torabi (2012c) also considered the DC method for the spatial models. Torabi (2013) studied the DC approach in the context of GLMM with measurement error in covariates.

The DC approach, thus, is well suited to address the issues in spatio-temporal analysis using the frequentist paradigm. Because these estimators are ML estimators, unlike the Bayesian estimators, they are independent of the choice of priors, non-estimable parameters are flagged automatically and possibility of improper posterior distribution is completely avoided (Lele *et al.* 2010).

In this paper, we use DC in the context of spatio-temporal analysis. In our spatio-temporal model, the well-known CAR model (Besag *et al.*, 1991) and B-splines are used for the spatial and temporal effects, respectively (Section 2). We, then, describe how DC can be used to obtain ML estimates, and also to get prediction and prediction intervals for

smoothing disease ratios over space and time (Section 3). In Section 4, the performance of the proposed approach is evaluated using a real dataset of yearly number of childhood asthma visits to hospital in the province of Manitoba, Canada, during 2000–2010. The performance of the DC approach (ML estimates) is also studied through a simulation study. Concluding remarks are given in Section 5.

2. SPATIO-TEMPORAL MODEL

Let y_{it} be the number of disease cases (or otherwise) for the *i-th* geographic area at time t, and let e_{it} be the corresponding expected number of disease cases for $i = 1, ..., I$; $t = 1, ..., T$, where e_{it} is typically calculated using overall rates across time and area (sometimes specific to age, race, and sex groups). Define $y_{it} \sim \text{Poi}(\mu^c_{it})$ where $\text{Poi}(\mu^c_{it})$ is a conditionally Poisson variable with mean function μ^c_{it} :

$$\mu^c_{it} = \exp\{\log e_{it} + m + S(t) + \eta_i + \theta_{it}\}, \tag{1}$$

where m reflects the overall ratio of the number of observed cases across the region to the number of cases expected across the region under the overall rates defining the e_{it}, η_i represents spatial pattern in the disease at region i, and θ_{it} is the interaction between the spatial and temporal effects. To account for the fixed temporal effects, $S(t)$ represents a cubic B-spline with one inner knot (Eilers, 1996). One may simply consider a linear trend (βt) instead of $S(t)$ depending on the nature of dataset. With the overall mean of ratio m in our model, the B-spline is provided without an intercept. In this case, $S(t)$ 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, 1996; De Boor, 2001). One

may use the AR model instead of B-spline, however, the B-spline model is more flexible than the AR model without also assuming any distribution. The CAR model is used to capture the spatial random effects η_i. A variety of CAR models may also 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 refers a set of neighbors for the i-th region (Besag *et al.*, 1991). We consider the following general model for the spatial effects η_i,

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

$$\Sigma_\eta = \sigma_\eta^2 (N_I - \lambda_\eta D)^{-1} P,$$

where P is a $I \times I$ diagonal matrix with elements $P_{ii} = 1/e_i$ with $e_i = \sum_{t=1}^{T} e_{it}$; D is a $I \times I$ matrix with elements $D_{ij} = (e_j/e_i)^{1/2}$ if region i and j are adjacent and $D_{ij} = 0$ otherwise; σ_η^2 is the spatial dispersion parameter; λ_η measures the conditional spatial dependence, $\lambda_{min} \leq \lambda_\eta \leq \lambda_{max}$, where $\lambda_{\min}^{-1}$ and λ_{max}^{-1} are the smallest and largest eigenvalues of $P^{-1/2} D P^{1/2}$; and N_I is an identity matrix of dimension I (Cressie and Chan, 1989; Stern and Cressie, 1999). One may define the interaction effect of space and time, θ_{it}, as $S_i(t), \delta_i t$, or simply iid Normal distribution as $\theta_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\theta^2)$, depending on the nature of dataset (Bernardinelli *et al.*, 1992; MacNab and Dean, 2001; Silva *et al.*, 2008; Torabi and Rosychuk, 2012). Note that δ_i is the coefficient of the linear temporal effect related to the i-th region, and $S_i(t)$ is a cubic B-spline for specific region i.

3. FREQUENTIST INFERENCE USING DATA CLONING

Let $\boldsymbol{y} = (y_{11}, ..., y_{1T}, ..., y_{I1}, ..., y_{IT})'$ be the observed data vector and, conditionally on the random effects, $\boldsymbol{v} = (\eta_1, ..., \eta_I, \theta_{11}, ..., \theta_{IT})'$, assume that the elements of $\boldsymbol{y}$ are independent and drawn from a Poisson distribution with parameters $\boldsymbol{\alpha}_1 = (m, \beta_1, ..., \beta_4)$. It is also assumed that distribution for $\boldsymbol{v}$ depends on parameters $\boldsymbol{\alpha}_2$ which includes $\lambda_\eta, \sigma_\eta^2$ and

related parameter(s) from θ_{it}; noting that $\boldsymbol{v}$ can have any appropriate distribution. The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$, and predict the incidence ratios over space and time as a function of $\boldsymbol{v}$.

To illustrate the DC approach, we start with standard Bayesian approach to inference for hierarchical models. Denote $L(\boldsymbol{\alpha}; \boldsymbol{y})$ as likelihood of $\boldsymbol{\alpha}$ given $\boldsymbol{y}$ and $\pi(\boldsymbol{\alpha})$ as prior distribution on the parameter space. The posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y})}, \tag{3}$$

where $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. There are computational tools, Markov chain Monte Carlo (MCMC) algorithms, that facilitate generation of random variates from the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ without computing the integrals in the numerator or the denominator of (3)(Gilks *et al.*, 1996; Spiegelhalter *et al.*, 2004).

The DC method uses the Bayesian computational approach for frequentist purposes. To understand the idea in DC, imagine a hypothetical situation where the observations $\boldsymbol{y}$ is 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 posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(K)}$ is then given by

$$\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)}) = \frac{\{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K \pi(\boldsymbol{\alpha})}{C(\boldsymbol{y}^{(K)})}, \tag{4}$$

where $C(\boldsymbol{y}^{(K)}) = \int \{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K \pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. The expression $\{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K$ is the likelihood for K copies of the original data. Lele *et al.* (2007, 2010) showed that, for K large enough, $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$ converges to a multivariate Normal distribution with mean equal to the MLE of the model parameters and variance-covariance matrix equal to $1/K$ times the inverse of the Fisher information matrix for the MLE. Hence, this distribution is nearly degenerated at the MLE $\boldsymbol{\alpha}$ for large K (Walker, 1969). Moreover, the sample

mean vector of the generated random numbers from (4) provides the MLE of the model parameters, and K times their sample variance-covariance matrix is an estimate of the asymptotic variance-covariance matrix for the MLE $\hat{\boldsymbol{\alpha}}$.

Lele *et al.* (2010) also provided various checks to determine the adequate number of clones K. For instance, one may plot the largest eigenvalue of the posterior variance as a function of the number of clones K to determine if the posterior distribution has become nearly degenerate. As another criterion, it is approximately true that as we increase the number of clones K,

$$(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' \boldsymbol{V}^{-1} (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) \sim \chi^2_p, \tag{5}$$

where $\boldsymbol{V}$ is the variance of the posterior distribution and p is the dimension of $\boldsymbol{\alpha}$. One may also compute the following two statistics: a) $\zeta = \frac{1}{B} \sum_{q=1}^{B} (O_q - Q_q)^2$, where O_q and Q_q are observed and estimated quantiles for χ^2_p random variable, and b) $\tilde{r}^2 = 1 - \rho^2$, where ρ is the correlation between (O_q, Q_q). If these statistics are close to zero, it indicates that the approximation (5) is reasonable.

3.1. Prediction of disease ratios

Prediction of random effects (disease ratios in our set up), particularly from the frequentist viewpoint, is usually problematic. If the parameters $\boldsymbol{\alpha}$ are known, then one can clearly use the conditional distribution of $\boldsymbol{DR} = (DR_{11}, ..., DR_{1T}, ..., DR_{I1}, ..., DR_{IT})'$, the latent variables, given the observed data; noting that $DR_{it} = \mu^c_{it}/e_{it}$ is the disease ratio at area i and time t. That is, one can use $\pi(\boldsymbol{DR}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$ where $\boldsymbol{\alpha}^*$ is the true value of the parameter. A naive approach, when $\boldsymbol{\alpha}$ is estimated using the data, is to use $\pi(\boldsymbol{DR}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}})$. However, this approach does not take into account the variability introduced by the model parameters estimate. An approach that has been suggested in the literature (e.g., Hamilton, 1986; Lele *et al.*, 2010) to take into account the variation of the estimator is

to use the density:

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

where $f(\cdot)$ and $g(\cdot)$ are Poisson and Normal distributions, respectively, and $\phi(., \mu, \sigma^2)$ denotes Normal density with mean μ and variance σ^2, which are equal to the MLE and the inverse of the Fisher information matrix here. In this paper, we obtain prediction (and prediction interval) of the $\boldsymbol{DR}$ using the density in equation (6) along with MCMC sampling. Note that we can use the same approach to predict, for example, $\exp(\eta_i + \theta_{it}), (i = 1, ..., I; t = 1, ..., T)$.

4. APPLICATION

4.1. Data analysis

We use a yearly dataset of childhood (age $\leq$ 20 years) asthma visits to hospital in the Canadian province of Manitoba during the 2000-2010 fiscal years. The population of Manitoba was stable during the study period from 1.15 million in 2000 to 1.20 million in 2010, with an average population of children of around 335,000. 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 (RHAD) and these RHAD are the geographic unit used in our model and all data were linked to these geographic boundaries. For simplicity, we call these regions 1,2,...,56. The number of childhood asthma visits totaled 14,690 over the study period with mean and median number of yearly cases per region of 26 and 17 (range 3 to 422), respectively. The regional child population sizes varied from 290 to 175,300, with mean and median numbers of 5,998 and 2,488, respectively. The largest population

was in region 56, while region 42 had the smallest population.

We first fit the model (1) to the dataset of children asthma cases using the DC, hierarchical Bayesian (HB), and PQL methods; noting that the expected number of asthma cases (e_{it}) was adjusted by gender. In particular, we have $e_{it} = \sum_{j=1}^{2} n_{itj} y_j / n_j$ where n_{itj} is the population at risk for the $i-$th geographic area, time t, and gender j; n_j is the population at risk for the gender j, and similarly y_j is the number of disease cases for the gender j. We used the PQL approach in our data analysis since this approach has been extensively used in the spatio-temporal context as a frequentist approach. Table 1 reports the model parameters estimate and corresponding standard errors for all three approaches; noting that we used $\theta_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\theta^2)$ which was found useful in our exploration of the data. It seems that the standard errors for some model parameters in DC approach are smaller than other two approaches; noting that the PQL based estimate for spatial random effect (σ_η^2) is so high compared to the DC and HB approaches.

"Table 1 around here"

In this paper, for the DC and HB analysis, the independent Normal distribution is assigned for fixed effects with zero mean and variance 10^6 and gamma distribution for the inverse of variance component with shape and scale parameter 0.001. Since the DC is invariant to the priors, one may use different priors. To monitor the convergence of the model parameters, we used several diagnostic methods implemented in the Bayesian output analysis (BOA) program (Smith, 2007), a freely available package created for R (R Development Core Team, 2012). We also used diagnostic methods implemented in the dclone package (Sólymos, 2010), which were described in Section 3, to monitor the convergence of the model parameters in terms of number of clones K (Lele *et al.* 2010). For this application, the number of clones was $K = 50$ to obtain MLE, and the number of iterations for convergence of the model parameters in DC was about 20,000. As mentioned

in Section 3, if scaled variances are decreasing at a $1/K$ rate and have reached a lower bound (say < 0.05), the DC approach has converged (Figure 1).

Other advantage of DC approach is that the non-estimable parameters are flagged automatically (Lele *et al.* 2010). To show this fact, we considered the model (1) with incorporating an iid regional random variable with Normal distribution (say, $\delta_i; i = 1, ..., 56$) and then fitted it to our dataset and observed that with increasing number of clones, the variances of posterior distributions also increase (not shown here).

"Figure 1 around here"

For a diagnostic analysis, we calculated 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

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

Figure 2 gives the residuals versus log-predicted diagnostic plot based on DC approach. It is clear from Figure 2 that there is no serious lack of fit in our model; noting that those 10 observations with relatively large log-predicted values in Figure 2 belong to Winnipeg health region (region 56, largest population).

"Figure 2 around here"

We now consider the provincial ratio of children asthma visits over time. Figure 3 shows the overall crude ratios of children hospital visits over time, $\sum_{i=1}^{I} y_{it} / \sum_{i=1}^{I} e_{it}$, and estimated ratios $\exp(m + \sum_{k=1}^{4} \beta_k B_k(t))$. It shows that over the study period, the cubic

B-spline produces smooth estimates of the crude ratios, and also an overall decrease in children asthma visits ratios over time exists.

"Figure 3 around here"

One of the main features of DC is the ability to predict the random effects. To have better understanding of the estimated spatial risk profile, we obtained the adjusted asthma ratio $\exp(\eta_i + \theta_{it})$ using DC, which provides a spatial risk profile. Figure 4 presents maps of the estimated spatial effects based on the fitted model, where the regional risk factor of asthma cases corresponds to some selected years. The overall spatial pattern suggests that some regions in the south and many regions 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. More investigation may be needed to explore the reasons for seemingly higher asthma ratios in these regions compared to other parts of the province.

"Figure 4 around here"

We also provide the regional asthma visits ratio estimate obtained from fitting the spatio-temporal mixed model given by $\exp(m + S(t) + \eta_i + \theta_{it})$. Figure 5 plots the fitted asthma visits ratio with corresponding 95% prediction intervals, for example, for health regions 37, 42, 52, and 56 using DC. The crude ratio estimates are y_{it}/e_{it}, and are also plotted in Figure 5. We indeed chose two regions with extreme population sizes; region 42 with least population and region 56 with largest population. As expected in Figure 5, our 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 a region would suggest that the underlying asthma visits rate in that region has also the same pattern relative to the provincial average.

"Figure 5 around here"

4.2. Simulation study

We conducted a simulation study to evaluate the performance of ML estimates, via DC approach, and compare with PQL and HB methods using a scenario similar to our children asthma dataset. More specifically, data are generated from the model (1) with the parameters ($m, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_\eta^2, \lambda_\eta$, and σ_θ^2) as listed in Table 2. The neighborhood structure and the population sizes are exactly as for the asthma dataset. Estimates for methods DC, PQL, and HB approaches are obtained with analyzing $1,000$ datasets generated from the mixed Poisson model (1).

Table 2 presents the bias values of the model parameters estimate, as well as the standard deviation of the estimated parameters and mean values of the estimated standard errors. In DC, the estimates are fairly unbiased and it seems that their standard errors are estimated reasonably well; noting that the conditional spatial dependence parameter (λ_η) had higher standard error and the spatial random effect (σ_η^2) was most biased in our simulation set-up which may indicate of underestimation of model parameters related to spatial random effects. The estimates in HB approach are also comparable with DC method unlike the PQL approach. Overall, it seems that DC approach, which provides a Monte carlo estimate of the MLE, gives good point estimates and standard errors for this data analysis.

"Table 2 around here"

5. Conclusion

Often, for fitting complex models in the spatio-temporal context, Bayesian methods are advocated because they are computationally more convenient than the likelihood-based methods. Analysis based on DC overcomes the computational difficulties of the ML method.

Using DC, we have implemented a model for spatio-temporal analysis that focused on the mapping of area level disease ratios (or rates). The model accommodated a CAR model for the spatial random effects and B-spline smoothing over the temporal effects. We proposed the DC method which provides a Monte carlo estimate of the MLE to estimate the model parameters, and also to provide prediction (and prediction intervals) of the smoothing incidence ratios over space and time. As another advantage of DC approach is that the non-estimable parameters are flagged automatically (Lele *et al.* 2010). In particular, we considered the model (1) with incorporating an iid regional random variable with Normal distribution and then fitted it to the dataset in Section 4.1 and observed that with increasing number of clones, the variances of posterior distributions also increase.

We adjusted our expected number of asthma cases by an important factor of gender. The model can be also easily extended to include some covariates directly. For instance, if the data are available, one may also want to consider other covariates such as exposure to dust and pollution, family history, obesity and so on.

Overall, it was suggested by the model estimates that the high asthma incidence ratios for children 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.

Our proposed approach to use DC is very general in the context of spatio-temporal models. In this paper, we used CAR and B-spline models for spatial and temporal effects,

respectively, however, one may consider other variants of spatial and temporal effects; for example CAR model (MacNab and Dean, 2001) for spatial random effects and AR models (Torabi and Rosychuk, 2010) for temporal random effects. We have planned to study model selection in the context of spatio-temporal models in the sense of frequentist approach using DC (Ponciano *et al.*, 2009).

Acknowledgements

I would like to thank two referees and an Associate Editor for constructive comments and suggestions which led to improvement of the manuscript. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
Disclaimer: The interpretations, conclusions and opinions expressed in this paper are those of the author and do not necessarily reflect the position of Manitoba Health. This study is based in part on data provided by Manitoba Health through Manitoba Centre for Health Policy. The interpretation and conclusions contained herein are those of the researcher and do not necessarily represent the views of the government of Manitoba.

REFERENCES

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

Besag J, York J, Mollià A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**:1–21.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *Journal of American Statistical Association* **88**:9–25.

Clayton DG, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**:671–681.

Cressie NA, Chan NH. 1989. Spatial modeling of regional variables. *Journal of the American Statistical Association* **84**: 393–401.

De Boor C. 2001. *A Practical Guide to Splines*, New York: Springer-Verlag.

Efron B. 1986. Why isn't everyone a Bayesian? *American Statistician* **40**:1–5.

Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* **11**:89–121.

Gilks WR, Richardson S, Spiegelhalter DJ. (eds.). 1996. *Markov Chain Monte Carlo in Practice*, New York: Springer - Verlag.

Hamilton JD. 1986. A standard error for the estimated state vector of a state-space model. *Journal of Econometrics* **33**:387–397.

Jonson NEG. 1991. Everyday diagnostics - A critique of the Bayesian model. *Medical Hypotheses* **34**:289–295.

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

Lele SR, Dennis B, 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.

Lele SR, Nadeem K, Schmuland B. 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association* **105**:1617–1625.

Liang KY, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* **73**:13–22.

MacNab YC, Dean CB. 2001. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics* **57**:949–956.

Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. 2008. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* **27**:2874–2889.

McCullagh P, Nelder JA. 1989. *Generalized Linear Models*, 2nd ed., London: Chapman and Hall.

McCulloch CE, Searle SR. 2001. *Generalized, Linear, and Mixed Models.* New York: John Wiley and Sons.

Ponciano JM, Taper ML, Dennis B, Lele SR. 2009. Hierarchical models in ecology: Confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* **90**:356–362.

Prentice RL, Zhao LP. 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**:825–839.

R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria ISBN 3-900051-07-0,http://www.R-project.org.

Silva GL, Dean CB, Niyonsenga T, Vanasse A. 2008. Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine* **27**:2381–2401.

Smith BJ. 2007. *BOA User Manual (version 1.1.7)*, Department of Biostatistics, College of Public Health, University of Iowa, Ames.

Sólymos P. 2010. dclone: data cloning in R. *The R Journal* **2**:29–37.

Spiegelhalter D, Thomas A, Best N, Lunn D. 2004. *WinBUGS version 1.4 User Manual.* MRC Biostatistics unit, Institute of Public Health,

Stern HS, Cressie NA. 1999. Inference for extremes in disease mapping. In Disease mapping and risk assessment for public health. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J-F. Viel and R. Bertollini (eds.), Chichester: Wiley, 63–84.

Torabi M. 2012a. Hierarchical Bayes estimation of spatial statistics for rates. *Journal of Statistical Planning and Inference* **142**:358–365.

Torabi M. 2012b. Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning. *Journal of Computational Statistics and Data Analysis* **56**:4259–4265.

Torabi M. 2012c. Spatial modeling using frequentist approach for disease mapping. *Journal of Applied Statistics*, doi: 10.1080/02664763.2012.711814.

Torabi M. 2013. Likelihood inference in generalized linear mixed measurement error models. *Journal of Computational Statistics and Data Analysis* **57**:549-557.

Torabi M, Rosychuk RJ. 2010. Spatio-temporal modelling of disease mapping of rates. *The Canadian Journal of Statistics* **38**:698–715.

Torabi M, Rosychuk RJ. 2011. Spatio-temporal modeling using spline for disease mapping: Analysis of childhood cancer trends. *Journal of Applied Statistics* **38**:1769–1781.

Torabi M, Rosychuk RJ. 2012. Hierarchical Bayesian spatiotemporal analysis of childhood cancer trends. *Journal of Geographical Analysis* **44**:109–120.

Walker AM. 1969. On the asymptotic behavior of posterior distributions. *Journal of Royal Statistical Society, Series B* **31**:80–88.

Waller LA, Carlin BP, Xia H, Gelfand AE. 1997. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**:607–617.

Zeger SL. 1988. A regression model for time series of counts. *Biometrika* **75**:621–629.

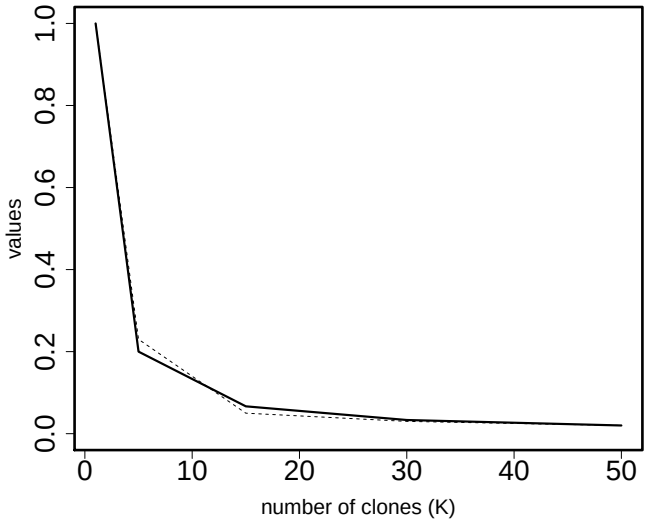


Figure 1. Data cloning convergence diagnostics for asthma visits to hospital dataset. The standardized maximum eigenvalues (solid line) converge to zero at the expected rate $1/K$ (dashed line)

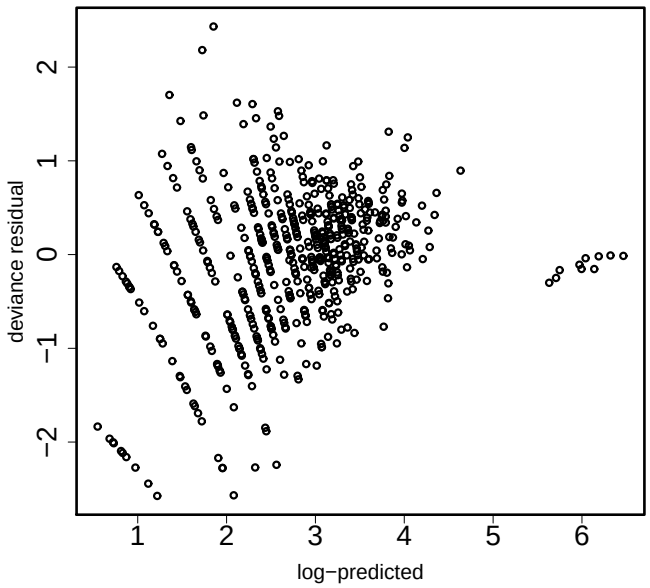
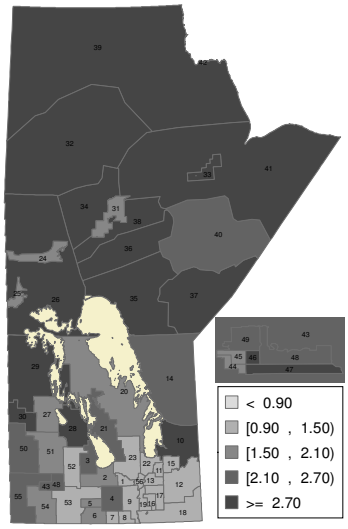
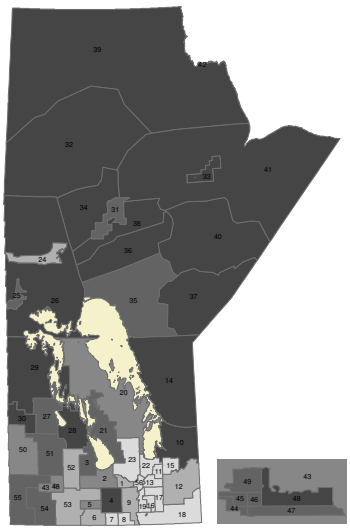


Figure 2. The deviance residuals versus log-predicted diagnostic plot of childhood asthma visits dataset based on data cloning approach

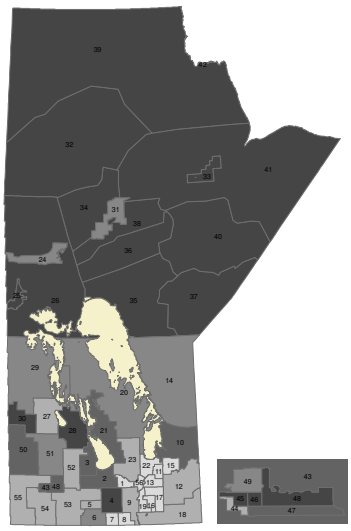
Figure 3. Provincial childhood asthma visits ratios over time. The solid black line represents cubic B-spline with blue and red lines as 95% confidence intervals; the dashed line is crude ratios



2001



2006



2010

Figure 4. Adjusted childhood asthma ratios for the spatial effects of the regional asthma risks for some selected years; Manitoba childhood asthma data (2000–2010). Major urban centre (Winnipeg region) is provided as inset

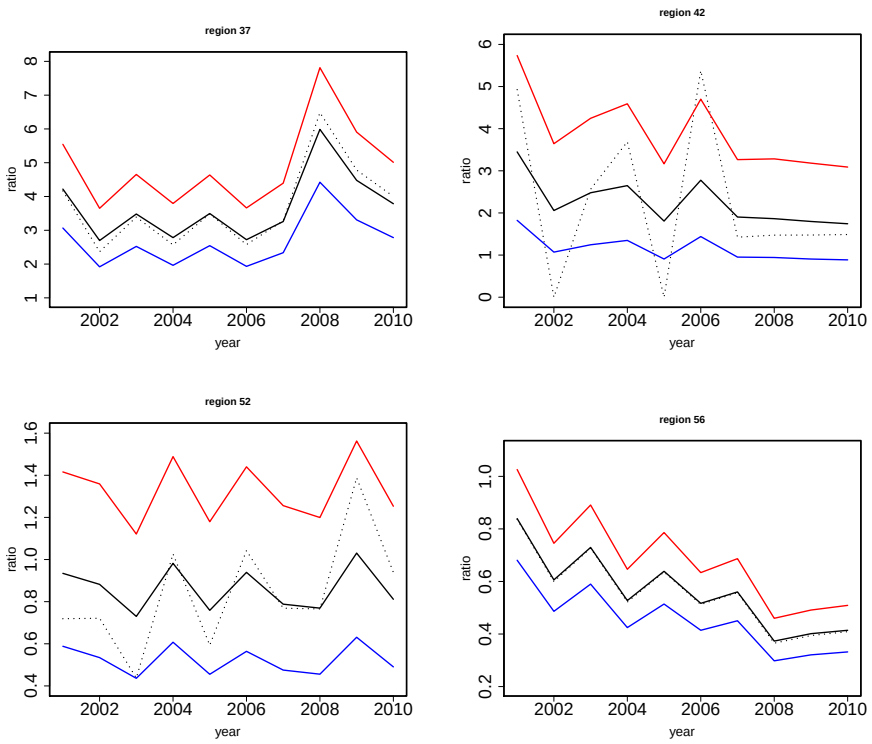


Figure 5. Fitted childhood asthma visits ratio to hospital for selected regions, 37, 42, 52, and 56 with yearly average population sizes 3821, 319, 2864, and 173439, respectively, during 2000–2010. The solid black line represents fitted ratios with blue and red lines as 95% prediction bands; the dashed line is crude ratios

Table 1. Parameter estimates (and standard errors), spatio-temporal mixed model for maximum likelihood estimation via data cloning (DC-MLE), penalized quasi-likelihood (PQL), and hierarchical Bayes (HB) methods, childhood asthma visits to hospital in the province of Manitoba, during 2000–2010

Parameter	Estimate (Standard error)		
	DC-MLE	PQL	HB
m	-1.210(0.810)	-1.184(0.890)	-1.195(0.930)
β_1	-0.296(0.088)	-0.117(0.088)	-0.446(0.100)
β_2	-0.397(0.130)	-0.271(0.129)	-0.506(0.140)
β_3	-0.593(0.120)	-0.405(0.114)	-0.737(0.124)
β_4	-0.662(0.067)	-0.493(0.066)	-0.798(0.074)
σ^2_η	0.336(0.072)	5.893(1e-7)	0.355(0.080)
λ_η	0.177(0.005)	0.281(0.0002)	0.173(0.007)
σ^2_θ	0.083(0.010)	0.075(0.008)	0.095(0.012)

Table 2. Mean values of biases and standard errors, and simulated standard errors of maximum likelihood estimation via data cloning (DC-MLE), penalized quasi-likelihood (PQL), and hierarchical Bayes (HB) methods based on 1,000 simulated datasets

Parameter	DC-MLE			PQL			HB		
	Bias	Standard error		Bias	Standard error		Bias	Standard error	
		DC-MLE	Simulated		PQL	Simulated		HB	Simulated
m =-1.20	0.001	0.814	0.791	0.010	0.851	0.810	0.001	0.805	0.780
β_1 =-0.10	-0.004	0.104	0.105	0.024	0.079	0.009	-0.001	0.104	0.106
β_2 =-0.25	0.001	0.151	0.158	0.147	0.138	0.012	0.001	0.152	0.158
β_3 =-0.40	-0.004	0.135	0.134	0.049	0.117	0.008	-0.004	0.135	0.135
β_4 =-0.50	0.001	0.079	0.078	-0.030	0.057	0.004	0.001	0.079	0.078
σ^2_η =1.00	-0.126	0.552	0.538	-0.923	0.116	0.035	-0.101	0.547	0.499
λ_η =0.01	-0.066	0.169	0.099	0.356	0.003	0.018	-0.063	0.127	0.054
σ^2_θ =0.08	-0.001	0.011	0.011	-0.006	0.010	0.0002	0.0001	0.011	0.011