NISS

Bayesian Methodology for Spatio-Temporal Syndromic Surveillance

Jian Zou and Alan F. Karr
David Banks and Matthew Heaton
Gauri Datta
James Lynch
Francisco Vera

Technical Report 174
September, 2010

Bayesian Methodology for Spatio-Temporal Syndromic Surveillance

Jian Zou and Alan F. Karr[1]
David Banks and Matthew Heaton[2]
Gauri Datta[3]
James Lynch[4]
Francisco Vera[5]

September 1, 2010

Abstract

Early and accurate detection of outbreaks is one of the most important objectives of syndromic surveillance systems. We propose a general Bayesian framework for syndromic surveillance systems. The methodology incorporates Gaussian Markov random field (GMRF) and Spatio-Temporal conditional autoregressive (CAR) modeling. By contrast, most previous approaches have been based on only spatial or time series models. The model has appealing probabilistic representations as well as attractive statistical properties. Based on extensive simulation studies, the model is capable of capturing outbreaks rapidly, while still limiting false positives.

Keywords: Syndromic surveillance, spatial statistics, Markov random field, spatio-temporal, conditional autoregressive process

1 Introduction

Syndromic surveillance uses syndrome (a specific collection of clinical symptoms) data that are monitored as potential indicators of a disease outbreak. The objective is to detect outbreaks as rapidly as possible, while also minimizing the number of false positives. In addition to the emphasis on timeliness, a syndromic surveillance system must also be able to (i) incorporate situation-specific characteristics such as covariate information for certain diseases; (ii) accommodate the spatial and temporal dynamics of the disease; (iii) integrate data from multiple sources; and (iv) provide analysis and visualization tools to help detect unexpected patterns.

In this paper we focus on the early and accurate detection of outbreaks of diseases, which could be either contagious or noncontagious. In syndromic surveillance, there is no definitive diagnosis of an outbreak at the early stage. A crucial assumption is that the disease syndrome is identical to that of the background. For example, the initial symptoms of anthrax are indistinguishable

---

[1]National Institute of Statistical Sciences, Research Triangle Park, NC
[2]Department of Statistical Science, Duke University, Durham, NC
[3]Department of Statistics, University of Georgia, Athens, GA
[4]Department of Statistics, University of South Carolina, Columbia, SC
[5]Department of Statistics, Clemson University, Clemson, SC

from those of several other endemic diseases such as influenza. A central challenge, then, is that all patients have the same syndrome, but some belong to the background while others have the disease to be detected. This inability to distinguish is reflected by the additivity in (3.1): the data are counts of people with the syndrome, who cannot be classified as background or outbreak.

As discussed in Section 3.1, an outbreak differs from the background in two respects: there is a higher number of cases, and there is temporal and spatial structure that is absent from the background.

We introduce a flexible hierarchical Bayesian model that can accommodate both spatial effects and temporal dynamics in a unified framework. We assume that the spatial aspects arise from the proximity of various measurements taken in adjacent sites. The temporal aspect is a direct result of a plausible Markov structure. Our results suggest that the model behaves sensibly and may be useful in even more complicated settings, such as when there are multiple data streams. Accounting for spatio-temporal correlation improves the assessment of the impact of outbreak distributions, produces accurate maps of occurrence, and allows for good prediction performance.

This paper is organized as follows. In Section 2 we review the literature on syndromic surveillance and discuss some of the drawbacks of existing models and methodology. In Section 3 we introduce our new spatio-temporal methodology to syndromic surveillance, and describe some properties of the model. In Section 4 we present some numerical studies and results. Finally, in Section 5 we discuss possible improvements of our current methods and future research directions.

2   Background

There exists a plethora of surveillance methods in the literature. One of the methods most widely used by public health departments is the CUSUM chart. This method was first proposed by Page (1954) to detect small shifts in the mean of a process. There followed many variants in the areas of quality control and disease surveillance (Ogden and Lynch 1999, Rossi et al. 1999, Cowling et al. 2006, Fricker et al. 2008). However, the CUSUM technique is relatively slow to respond to large shifts. Also, special patterns, such as spatial and temporal dynamics, are hard to detect and analyze.

Spatial scan statistics, proposed by Kulldorff (1997), have been applied to a wide variety of epidemiological studies for disease cluster detection. However, this method suffers from several drawbacks. It has difficulty finding small clusters other than the primary cluster, it lacks measures of uncertainty associated with the identified cluster, and it is unable to account for covariate information.

From a different direction, there is a long history of Bayesian modeling of change point problems, including Bayesian analysis of changing linear models and time series models. This subject is also extensively studied in time series analysis and identification. Many practical problems arising in quality control, recognition-oriented signal processing, and fault detection and monitoring in industrial plants, can be modeled with the aid of parametric models in which the parameters are subject to abrupt changes at unknown time instants.

This topic can be divided into retrospective or off-line analysis of the change point problem and sequential change point detection such as the Wald-type sequential probability ratio tests (Spiegelhalter, et al. 2003). While retrospective change point analysis typically focuses on estimation of when the change point occurred, sequential analysis is a hypothesis testing problem which can be formulated as a Bayesian model selection problem (Giron, Moreno and Casella 2007). Smith (1975) developed a Bayesian approach that was restricted to discrete time analysis, where the index of a

sequence of random variables corresponding to the change point was estimated. Carlin, Gelfand and Smith (1992) developed a Bayesian approach for change points having continuous support. They considered Bayesian analysis of a Poisson process with a single change point using Dirichlet priors on the transition matrix. Barry and Hartigan (1993) discussed alternative formulations of the change points in terms of product partition distributions. They assume that the observations are independently and normally distributed, and that the probability of a change point at a position i is p. The independence can be weakened, but some kind of conditional independence is required, for instance that observations in different blocks are mutually independent. McCulloch and Tsay (1993) give a Bayesian method for estimating random level and variance shifts in an autoregressive time series. Their method is based on estimation of the probability and size of a shift at each time point, and is a generalization of the usual odds ratio for model discrimination in Bayesian inference. Chib (1998) proposed models for multiple change points in which the change point probability is not constant but depends on the regime. Giron, Moreno and Casella (2007) proposed an objective Bayesian method for multiple change points in linear models based on the intrinsic priors.

Change point models are not directly applicable to our setting of syndromic surveillance because changes occur at different times in different locations. Moreover, none of these methods accounts for spatial or temporal correlation in the data, which can lead to higher prediction errors and underestimated standard errors in the inference results.

Bayesian hierarchical models have become increasingly popular in the analysis of spatial and spatio-temporal data (Banerjee, Carlin and Gelfand 2004). The advantage of such models is that they can make use of available prior information while simultaneously borrowing strength from the data when estimating the quantities of interest. They also provide explicit quantification of uncertainties, which is essential in real applications. Numerous models have been developed for disease mapping and the spread of existing outbreaks. To model spatial similarity in a conditional fashion, Besag (1974, 1975) pioneered the use of conditional autoregressive (CAR) models, whose computational convenience motivates much of the recent disease mapping literature (Clayton and Kaldor 1987; Besag, York and Mollie 1991; Besag and Kooperberg 1995; Waller et al. 1997 and Sun et al. 2000).

To date, there have been several Bayesian treatments in the area of syndromic surveillance. For example, Banks et al. (2010) proposed to accommodate spatial variations using a conditional probabilistic approach. They used the CAR model to account for spatial dependence among the locations of the drug abuse reporting centers. Similar model-based approach has been considered in Knorr-Held and Richardson (2003), Martínez-Beneito et al. (2008), and Zhou and Lawson (2008).

3   Methods

In this paper, we mainly focus on abrupt changes happening in discrete time and on contagious diseases.

3.1   Model

The basic model described in Banks et al. (2010) is adopted here. Specifically, let $Y_i(t)$ be the number of individuals with a specific syndrome recorded at hospital i on day t, where $i = 1, \ldots, m$ and $t = 1, \ldots, T$. We assume that when a disease outbreak occurs, *both the level and the spatio-temporal structure* of the $Y_i(t)$ change.

Spatial relationships between hospitals are represented by an adjacency matrix $W = (w_{ij})$: if

hospitals i and j are adjacent, then $w_{ij} = 1$, and otherwise $w_{ij} = 0$. In the example in Section 4, the reporting units, rather than hospitals, are the 100 counties in the state of North Carolina, and adjacency means "share a common border." More complicated definitions, such as the "piecewise linear decay" in Green and Richardson (2002), can readily be accommodated, but may cause W no longer to be sparse, adding to the computational effort in the inference procedure.

We model the number of counts $Y_i(t)$ by a Bayesian hierarchical model. We assume the first stage is Poisson with canonical link (log linear), so that in the absence of an epidemic, the mean function of the Poisson count at hospital i is $\mu_i(t)$. When there is an epidemic, a second component is added (recall that counts are of people with a syndrome shared by the background and disease.) to the baseline. We use an indicator function $\delta_i(t)$ as the mark of whether the epidemic is present. The additional intensity in epidemic state is represented by $\lambda_i(t)$. Thus, the first stage model becomes

$$Y_i(t) \sim \text{Pois}\left(\mu_i(t) + \delta_i(t)\lambda_i(t)\right), \tag{3.1}$$

where we assume $\boldsymbol{\mu}$, $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are mutually independent.

Model for $\boldsymbol{\mu}$. Let $\theta_i(t) = \log(\mu_i(t))$. We assume that $\theta_i(t) = \boldsymbol{X}_i^T(t)\boldsymbol{\beta}_\mu + \varepsilon_i(t)$, where $\boldsymbol{X}_i = (1, X_{i,1}, \cdots, X_{i,p})^T$, $i = 1, \ldots, m$, represent covariates such as population size, $\boldsymbol{\beta}_\mu = (\beta_{\mu,0}, \beta_{\mu,1}, \cdots, \beta_{\mu,p})^T$ are covariate coefficients, and $\varepsilon_i(t) \sim N(0, \sigma_\mu^2)$. Spatial and temporal variations can be incorporated in the covariates. Nevertheless, when $\boldsymbol{\beta}_\mu = 0$ as illustrated in Section 4, the background is devoid of both spatial and temporal structure.

Model for $\boldsymbol{\lambda}$. When there is an outbreak, we assume that the additional intensity $\lambda_i(t)$ follows a model with spatio-temporal structure. Specifically, let $\eta_i(t) = \log(\lambda_i(t))$; then,

$$\eta_i(t) = \boldsymbol{U}_i^T(t)\boldsymbol{\beta}_\lambda + \xi_i(t),$$

where $\boldsymbol{\beta}_\lambda = (\beta_{\lambda,0}, \beta_{\lambda,1}, \cdots, \beta_{\lambda,q})^T$ are covariate coefficients. We assume that the first column of $\boldsymbol{U}_i$ consists entirely of ones, in which case $\beta_{\lambda,0}$ becomes a scaling factor that can be interpreted as the relative size of the outbreak compared to the baseline. The $\xi_i(t)$ are stipulated to satisfy

$$\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1) \sim N\left(\frac{\rho_1}{w_{i+}}\sum_{j \neq i} w_{ij}\xi_j(t) + \rho_2\xi_i(t-1), \frac{\sigma_\lambda^2}{w_{i+}}\right). \tag{3.2}$$

In (3.2), $\boldsymbol{\xi}_{-i}(t) = (\xi_1, \cdots, \xi_{i-1}, \xi_{i+1}, \cdots, \xi_m)^T$, $w_{i+} = \sum_{j \neq i} w_{ij}$, ρ_1 is a spatial correlation and ρ_2 is a temporal correlation. We take $\boldsymbol{\xi}(1) = (0, \cdots, 0)^T$ as the initial values at $t = 1$.

Since our goal is to detect outbreaks as early as possible, it is of primary interest to study scenarios when the size of the additional counts is relatively small: $\boldsymbol{\lambda}$ is of the same order as $\boldsymbol{\mu}$. However, it is very hard to distinguish the "signal" $\boldsymbol{\lambda}$ from the "noise" $\boldsymbol{\mu}$ when the relative change $e^{\beta_{\lambda,0}}$ is too small, in which case the system may have too many false alarms. This issue motivates our proposing different model structures for $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. By assuming white noise background and spatio-temporal heterogeneity in the signal, we increase the power to detect.

Model for $\boldsymbol{\delta}$. Let $\delta_i(t) = 1$ if the disease is present at hospital i on day t and $\delta_i(t) = 0$ otherwise. Currently, we employ an *absorbing state model* for $\boldsymbol{\delta}$:

$$P\left(\delta_i(t+1) = 1|\boldsymbol{\delta}(t)\right) = \begin{cases} 1 & \text{if } \delta_i(t) = 1, \\ p_s\mathbf{1}(\delta_j(t) = 0 \,\forall j \in N_i) + 1 - (1 - p_c)^{\sum_{j \in N_i} \delta_j(t)} & \text{if } \delta_i(t) = 0. \end{cases} \tag{3.3}$$

where N_i is the set of spatial neighbors of i, that is, $N_i = \{j : W_{ij} = 1\}$. We assume that the $\delta_i(t+1)$ are conditionally independent given $\boldsymbol{\delta}(t)$. The two parameters in (3.3) have straightforward

interpretations: p_s is spontaneous generation rate for outbreaks, i.e., the probability of an outbreak when neither the site nor any neighbors has an outbreak, and p_c is contagion rate for transfer of outbreaks at neighbors to a site without an outbreak.

Notice that the two terms involving p_s and p_c represent mutually exclusive behaviors. That is, if the outbreak is absent at all neighbors of i at time t, then it is generated spontaneously at time $t+1$ with probability p_s; or if the outbreak is present at any of the neighbors of i then the probability of location i being infected at time $t+1$ is $1 - (1 - p_c)^{\sum_{j \in N_i} \delta_j(t)}$.

3.2 Properties

In this section, we look at some properties of the model given in (3.2).
First, recall the model for the spatio-temporal random effect

$$\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1) \sim N\left(\frac{\rho_1}{w_{i+}}\sum_{j \neq i} w_{ij}\xi_j(t) + \rho_2\xi_i(t-1), \frac{\sigma_\lambda^2}{w_{i+}}\right) \tag{3.4}$$

From (3.4), we have the conditional mean of the form

$$E\left[\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1)\right] = \sum_{j \neq i} \rho_1 \frac{w_{ij}}{w_{i+}} \xi_j(t) + e_i, \tag{3.5}$$

where $e_i = \rho_2 \xi_i(t-1)$. Next we need to find the distribution (i.e., the mean vector and covariance matrix) of $(\xi_1(t), \cdots, \xi_m(t))^T$ given $\boldsymbol{\xi}(1), \cdots, \boldsymbol{\xi}(t-1)$. Let

$$\boldsymbol{\nu}(t) = E\left[\boldsymbol{\xi}(t)|\boldsymbol{\xi}(1), \cdots, \boldsymbol{\xi}(t-1)\right].$$

Then from (3.5) we get

$$\nu_i(t) = \sum_{j \neq i} \rho_1 \frac{w_{ij}}{w_{i+}} \nu_j(t) + e_i, \quad i = 1, \ldots, m, \tag{3.6}$$

so that

$$\boldsymbol{\nu}(t) = \rho_1 \tilde{W} \boldsymbol{\nu}(t) + \boldsymbol{e}, \tag{3.7}$$

where $\boldsymbol{e} = (e_1, \cdots, e_m)^T$ and $\tilde{W} = (\tilde{w}_{ij}) = (w_{ij}/w_{i+})$ is the scaled adjacency matrix. That is,

$$\tilde{B}\boldsymbol{\nu}(t) = \boldsymbol{e}, \tag{3.8}$$

where $\tilde{B} = I - \rho_1 \tilde{W}$ has diagonal elements 1 and off-diagonal elements $-\rho_1 \tilde{w}_{ij}$. If $\tilde{B}$ is nonsingular, then

$$\boldsymbol{\nu}(t) = \tilde{B}^{-1}\boldsymbol{e} = \rho_2 \tilde{B}^{-1}\boldsymbol{\xi}(t-1). \tag{3.9}$$

Using (3.8), we can rewrite (3.5) as

$$E\left[\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1)\right] = \nu_i(t) + \sum_{j \neq i} \rho_1 \tilde{w}_{ij}\left(\xi_j(t) - \nu_j(t)\right). \tag{3.10}$$

Also we have conditional variances

$$\text{Var}\left[\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1)\right] = \frac{\sigma_\lambda^2}{w_{i+}} = \sigma_i^2 \text{ (say)}. \tag{3.11}$$

Then after simplification we get

$$\text{Var}\left[\boldsymbol{\xi}(t)|\boldsymbol{\xi}(t-1)\right] = \tilde{B}^{-1}D = [I - \rho_1\tilde{W}]^{-1}D,$$

where $D = \text{Diag}(\sigma_i^2)$. Here note that in order to have a symmetric covariance matrix, we specify $\tilde{W} = \text{Diag}(1/w_{i+})W$. Thus,

$$\boldsymbol{\xi}(t)|\boldsymbol{\xi}(1),\ldots,\boldsymbol{\xi}(t-1) \sim \text{MVN}\left(\rho_2\tilde{B}^{-1}\boldsymbol{\xi}(t-1), \tilde{B}^{-1}D\right). \tag{3.12}$$

With $D_w = \text{Diag}(w_{i+})$ and $\sigma_i^2 = \sigma_\lambda^2/w_{i+}$, then

$$\tilde{B}^{-1}D = (I - \rho_1 D_w^{-1}W)^{-1}D_w^{-1}\sigma_\lambda^2 = (D_w - \rho_1 W)^{-1}\sigma_\lambda^2. \tag{3.13}$$

Therefore, provided that $(D_w - \rho_1 W)$ is positive definite, the conditional joint distribution of $\boldsymbol{\xi}(t)$ given $\boldsymbol{\xi}(t-1)$ is

$$\boldsymbol{\xi}(t)|\boldsymbol{\xi}(t-1) \sim (2\pi)^{-m/2}|\tilde{B}^{-1}D|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\xi}(t) - \rho_2\tilde{B}^{-1}\boldsymbol{\xi}(t-1)\right)^T\right.$$

$$\left.(D_w - \rho_1 W)\left(\boldsymbol{\xi}(t) - \rho_2\tilde{B}^{-1}\boldsymbol{\xi}(t-1)\right)\right\}. \tag{3.14}$$

Next, we discuss two special cases. First, suppose that $\rho_2 = 0$, in which case (3.2) becomes a CAR model. We write $\tau = \Sigma_{\boldsymbol{\xi}}^{-1} = D_w - \rho_1 W$, that is, we choose ρ_1 to make τ nonsingular. According to Banerjee, Carlin and Gelfand (2005), τ is positive definite if $\rho_1 \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$ where $\lambda_{(1)}$ and $\lambda_{(n)}$ are the smallest and the largest eigenvalues of $D_w^{-1/2}WD_w^{-1/2}$. Note that the spatial neighborhood structure is determined by the precision matrix. For $i \neq j$, $(\Sigma_{\boldsymbol{\xi}}^{-1})_{ij} = 0$, or equivalently $w_{ij} = 0$, implies that $\xi_i(t)$ and $\xi_j(t)$ are conditionally independent given the other ξ's. This is referred to as the local Markov property with first-order neighbor structure, which expresses the independence of the vertices with its neighbors in a graph. If this property holds for all the vertices and its neighbors, then we have a Gaussian Markov random field (GMRF), and vice versa.

Another note about the network model is that we can easily incorporate weights on different locations and edges, but the precision matrix needs to be adjusted according to the weights. For example, we can take the weight between two neighbors to be inversely proportional to the distance between them. Then our $\tau_{ii} = w_{i+}^*/\sigma^2$ and $\tau_{ij} = w_{i+}^*/w_{ij}^*$ where W^* is the weighted adjacency matrix.

By contrast, suppose that $\rho_1 = 0$, in which case (3.2) becomes a VAR(1) model, which can also be expressed as

$$\boldsymbol{\xi}(t) = \Phi\boldsymbol{\xi}(t-1) + \boldsymbol{\epsilon}(t) \tag{3.15}$$

where $\Phi = \text{Diag}(\rho_2)$ for $-1 < \rho_2 < 1$, $t = 2,\ldots,n$ and $\epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2)$. Then the joint distribution of $\boldsymbol{\xi}$ is a multivariate normal with covariance matrix having the Toeplitz form

$$\Sigma = \frac{\sigma^2}{1-\rho_2^2}\begin{bmatrix} 1 & \rho_2 & \rho_2^2 & \cdots & \cdots & \rho_2^{n-1} \\ \rho_2 & 1 & \rho_2 & \ddots & & \vdots \\ \rho_2^2 & \rho_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \rho_2 & \rho_2^2 \\ \vdots & & \ddots & \rho_2 & 1 & \rho_2 \\ \rho_2^{n-1} & \cdots & \cdots & \rho_2^2 & \rho_2 & 1 \end{bmatrix}. \tag{3.16}$$

In this case, the precision matrix has a nice sparse structure as

$$\tau = \sigma^{-2} \begin{bmatrix} 1 & -\rho_2 & 0 & \dots & \dots & 0 \\ -\rho_2 & 1+\rho_2^2 & -\rho_2 & \ddots & & \vdots \\ 0 & -\rho_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\rho_2 & 0 \\ \vdots & & \ddots & -\rho_2 & 1+\rho_2^2 & -\rho_2 \\ 0 & \dots & \dots & 0 & -\rho_2 & 1 \end{bmatrix}. \tag{3.17}$$

The band along the diagonal results from the conditional independence of $\boldsymbol{\xi}(t_i)$ and $\boldsymbol{\xi}(t_j)$ for $|t_i - t_j| > 1$, given the rest.

Therefore, model (3.2) has the feature that if we look at either the space or time domain separately, we will have conditional independence across space or across time. The idea is similar to a mixture model; however, the relationship of the conditional independence and the zero structure of the precision matrix is not evident in the joint model. In addition, the conditional variance σ^2 must be small enough to ensure positive definiteness of the precision matrix. One sufficient condition is diagonal dominance, that is

$$\tau_{ii} = w_{i+}/\sigma^2 > \sum_{i \neq j} |\tau_{ij}|.$$

Finally, in the conditional specification (3.4), we could use two versions of the adjacency matrix, the 0-1 adjacency matrix W and the scaled adjacency matrix $\tilde{W}$. These two cases correspond to constant and nonconstant conditional variance in the model assumptions. Regarding $\tilde{W}$, there is an alternative approach that would instead use

$$\xi_i(t)|\boldsymbol{\xi}_{-i}(t), \boldsymbol{\xi}(t-1) \sim N\left(\rho_1 \sum_{j \neq i} w_{ij}\xi_j(t) + \rho_2\xi_i(t-1), \sigma^2\right). \tag{3.18}$$

The distinction between (3.4) and (3.18) is that the former tries to stabilize the conditional variance. The conditional mean at a site is additive, based on all its neighbors. On the other hand, (3.18) focuses on stabilizing the mean, that is, to make the mean shrink towards the average of its spatial neighbors. However, in this case, extra care should be given to avoid running into the risk of having the temporal component dominate the spatial term. Furthermore, the model given in (3.18) does not make so much sense in our setting.

4 Numerical Studies

In this section we present the results of initial simulation experiments. We employ simulations in order to evaluate our method by comparing its results to a known (albeit simulated) reality.

4.1 Prior Distributions for μ, δ and λ

As discussed previously, in these experiments we take the background process $\boldsymbol{\mu}(t)$ to be pure noise, with neither spatial nor temporal structure. An alternative is to assume that $\boldsymbol{\mu}(t)$ has a non-trivial spatial structure differing from that of $\boldsymbol{\lambda}(t)$. Hence, we can take advantage of the additional variability in the observations when the epidemic is present.

We assume independent prior distributions for $\boldsymbol{\mu}$, $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$, i.e.,

$$p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\lambda}) = p(\boldsymbol{\mu})p(\boldsymbol{\lambda})p(\boldsymbol{\delta}). \tag{4.1}$$

The specific forms of these priors are as follows. First,

$$p(\boldsymbol{\mu}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{m/2}|\Sigma_{\boldsymbol{\mu}}|^{-1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\mu}(t)^T \Sigma_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu}(t)\right) \tag{4.2}$$

where $\Sigma_{\boldsymbol{\mu}} = \text{Var}(\boldsymbol{\mu}) = \sigma^2 I$. The prior distribution $p(\boldsymbol{\lambda})$ has the form

$$p(\boldsymbol{\lambda}) = \prod_{t=2}^{T} \frac{1}{(2\pi\sigma^2)^{n/2}|B^{-1}|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\zeta}(t) - \rho_2 B^{-1}\boldsymbol{\zeta}(t-1)\right)^T\right.$$
$$\left. B\left(\boldsymbol{\zeta}(t) - \rho_2 B^{-1}\boldsymbol{\zeta}(t-1)\right)\right\}|J_2|, \tag{4.3}$$

where $\boldsymbol{\zeta} = \log(\boldsymbol{\lambda})$, $B = I - \rho_1 W$ and $J_2 = \text{Diag}(e^{-\zeta_i})$. Finally,

$$p(\boldsymbol{\delta}) = \prod_{t=2}^{T}\prod_{i=1}^{m} p_{it}^{\delta_i(t)}(1 - p_{it})^{1-\delta_i(t)}p_{i1}, \tag{4.4}$$

where $p_{it} = P\left(\delta_i(t+1) = 1|\boldsymbol{\delta}(t)\right)$ and $p_{i1} = p_s$.

Note that we cannot recover p_s and p_c exactly from one realization of the model: there is simply not enough information in the data for consistent estimation. Therefore, prior information must be available in order to conduct the inference. In reality, we know that p_s must be small, and p_c can be empirically estimated from historical data for known diseases.

Another interesting aspect is the effect of a different time scale or frequency of data collection. For example, if data were collected weekly instead of daily, the inference will differ. One possible solution is to slow down the process by decreasing the values of p_s and p_c.

Also, prediction errors are related to the estimation of the parameter p_s. For instance, suppose the estimate of the spontaneous generation rate for outbreaks p_s is $\hat{p}_s$, then $\hat{p}_s > p_s$ will likely result in predicting outbreaks too early, hence a type I error. On the other hand, $\hat{p}_s < p_s$ will likely result in predicting outbreaks too late, corresponding to a type II error.

The model is then completed with the prior specifications for the hyperparameters $(\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\lambda, \sigma_\mu, \sigma_\lambda, \rho_1, \rho_2, p_s, p_c)$:

$$\boldsymbol{\beta}_\mu \sim N(0, dI) \tag{4.5}$$
$$\boldsymbol{\beta}_\lambda \sim N(0, dI) \tag{4.6}$$
$$\sigma_\mu^2 \sim IG(a_1, b_1) \tag{4.7}$$
$$\sigma_\lambda^2 \sim IG(a_2, b_2) \tag{4.8}$$
$$\rho_1 \sim \text{Unif}(-1, 1) \tag{4.9}$$
$$\rho_2 \sim \text{Unif}(-1, 1) \tag{4.10}$$
$$p_s \sim \text{LogNormal}(c_1, d_1)I(0, 1) \tag{4.11}$$
$$p_c \sim \text{LogNormal}(c_2, d_2)I(0, 1). \tag{4.12}$$

The choices of hyperparameters represent vague prior information and ensure posterior propriety; see also discussions on the restrictions on some parameters for these hyperpriors in Sun et al. (2000).

All of the previous specifications are expressed in a conditional fashion. That is, each of the model components is actually a conditional distribution for the stated variable, conditional on both its immediate parameters and hyperparameters. A useful graphical tool for representing this hierarchical Bayesian model is the directed acyclic graph, or DAG. In Figure 1, the likelihood function is represented as the root of the graph; each prior is represented as a separate node pointing to the node that depends on it. By working through this diagram, which graphically characterizes the dependence structure between variables, one can identify possible causal effect and decide on appropriate analytic priors.
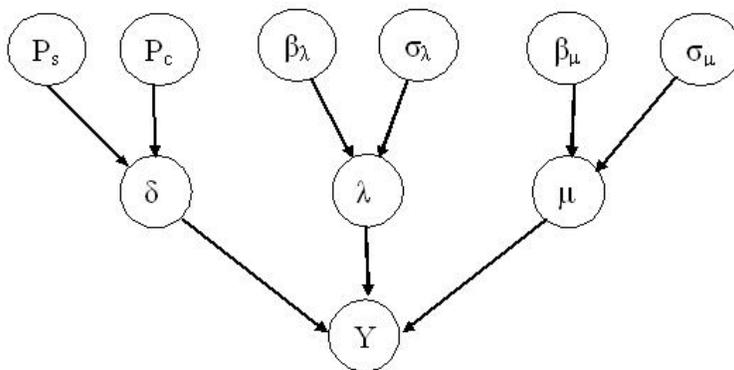
Figure 1: Directed acyclic graph for the full model

4.2   Simulation Methodology

As noted earlier, the setting is the 100 counties in the state of North Carolina, with $w_{ij} = 1$ if counties i and j are geographically adjacent, and $w_{ij} = 0$ otherwise. The counties and their numbers of neighbors are shown in Figure 2. This network has a very sparse adjacency matrix W, which makes sparse matrix computational techniques applicable.

It is also reasonable to assume that $\sigma_\lambda > \sigma_\mu$, as one may expect larger variability in an epidemic than in a non-epidemic phase. Therefore, we set $\sigma_\lambda = 0.5$ and $\sigma_\mu = 0.1$ in the data generation process. To mimic real scenarios, we generated data with parameters $p_s = 0.001$, $p_c = 0.2$, $\rho_1 = 0.5$, $\rho_2 = 0.5$, and $\beta_{\lambda,0} = 1$. To begin the simulation, we initialized the states at time $t = 1$ as $\boldsymbol{\xi(0)} = (0, \ldots, 0)^T$, $T = 22$, and $e^{\beta_{\mu,0}} = 5$. We avoid identifiability and posterior propriety issues by imposing vague proper CAR priors in the inference, We used two parallel chains started from different values, each run for 1000 iterations as burn-in for the process to reach stationarity, then the next 5000 iterations are kept for posterior analysis. Trace and autocorrelation plots did not reveal any convergence problems.

Model inferences are carried out via Markov chain Monte Carlo (MCMC) simulation that provides samples from the posterior distribution of all quantities of interest. Bayesian computation is conducted using Bayesian inference Using Gibbs Sampling (BUGS) software. Simulations and plots are done in R (CRAN). There is also a "R2WinBUGS" package available in R which provides
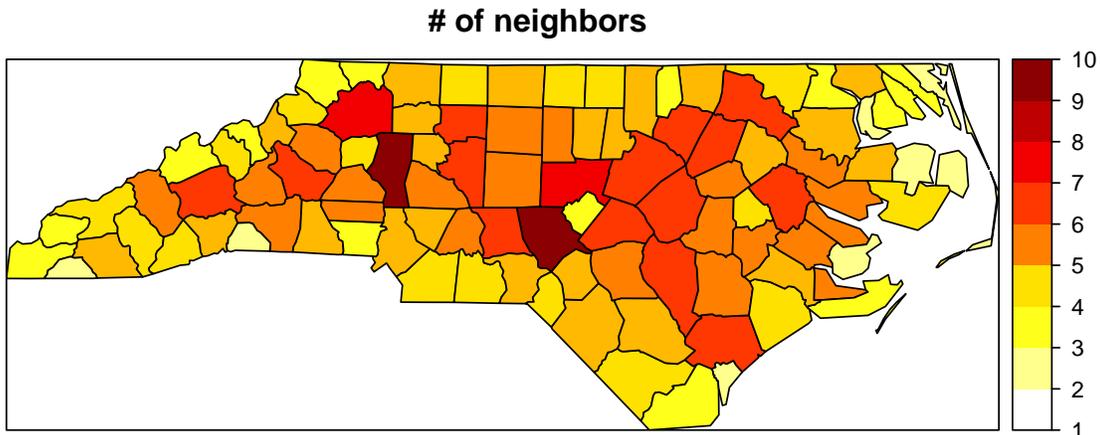
of neighbors

Figure 2: Distribution of number of neighbors for each county

convenient functions to call WinBUGS from R, summarize inferences and create tables and graphs. Since most of the conditionals are not in closed form, an adaptive MCMC algorithm (Haario et al. 2001) is utilized in the sampling procedure. The computation times vary depending on the length of the time series for the dataset. For our simulation, a complete run including both prospective and retrospective analyses takes several hours (See Section 4.3 for prospective and retrospective analyses). Computational time can be significantly reduced if the program is coded in a lower-level language, e.g., C or Fortran. Nonetheless, the current implementation is still acceptable when data are collected at daily time intervals.

4.3   Simulation Results

We applied our methods to simulated observations generated by means of model (3.1). We focused on two families of estimators. The first, $P(\boldsymbol{\delta}(t) = (\cdot)|Y_{1:t})$ where $t \leq T = 22$, are the "real time" estimators that would be used to attempt to detect an outbreak, using which $\boldsymbol{\delta}(t)$ is estimated from $Y(1), \ldots, Y(t)$. The second family of estimators $P(\boldsymbol{\delta}(t) = (\cdot)|Y_{1:T})$, represent *ex post facto* (retrospective) reconstruction of the epidemic, using the full data $Y(1), \ldots, Y(T)$ to estimate $\boldsymbol{\delta}(t)$ for each day t.

Finally, we analyzed 100 synthetic replicate datasets, and two cases are presented in Figures 3 and 4. In these figures, T stands for the true δ process, R stands for real-time probabilities $P(\delta_i(t) = 1|Y_{1:t})$, and E represents *ex post facto* probabilities $P(\delta_i(t) = 1|Y_{1:T})$.

From the retrospective analysis, the latent $\boldsymbol{\delta}$ process can be recovered almost perfectly. The posterior means of $P(\delta_i(t) = 1|Y_{1:T})$ shows the same spatial patterns as the underlying true $\boldsymbol{\delta}$ process as depicted in the third column of Figures 3 and 4. In Figure 3, two outbreaks occur simultaneously at time $T = 9$ and gradually propagate to their neighbors. The *ex post facto* probabilities for those two counties at $T = 9$ both exceed 0.75. The spatial and temporal dynamics are reflected with high precision in the subsequent plots. Another thing to note is that at the time point just prior to the start of the epidemic, the model gives some low probabilities in the two regions. This is a smoothing effect: the model has to trace back and determine the exact place and time of the outbreak.

10

The real-time estimators are, as they must be, less effective than the *ex post facto* ones. For instance in column R of Figure 4, there are many counties with medium probabilities of $\delta = 1$ on days 9 and 10; these could constitute false alarms if the threshold is too low. We hypothesize that this "anxiousness to declare an epidemic" results from the fact that the absorbing state model assumes an outbreak is going to occur. So the longer time goes without detecting an outbreak, the more anxious the absorbing state model becomes. However, from the numerical studies we found that once the epidemic is established, the real-time estimators track its spread accurately.
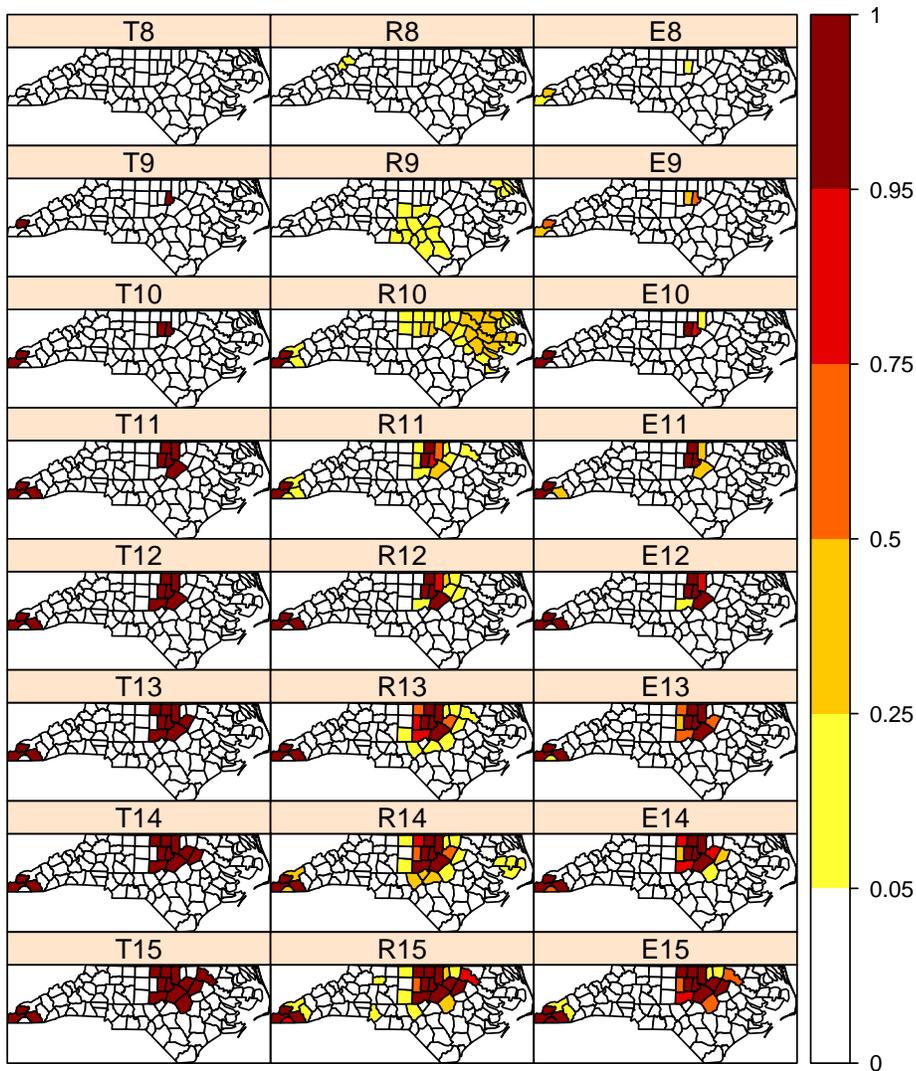
Figure 3: True vs Real time vs. *ex post facto* probabilities

Since there is always a tradeoff between quick detection and fewer false alarms, one must assess the performance of the surveillance system based on both criteria. A false alarm is defined when there is no outbreak but a prespecified threshold for $P(\delta_i(t) = 1|Y_{1:t})$ is exceeded. On the other hand, we define timeliness by the number of days delayed in detecting an outbreak. Figure 5 shows
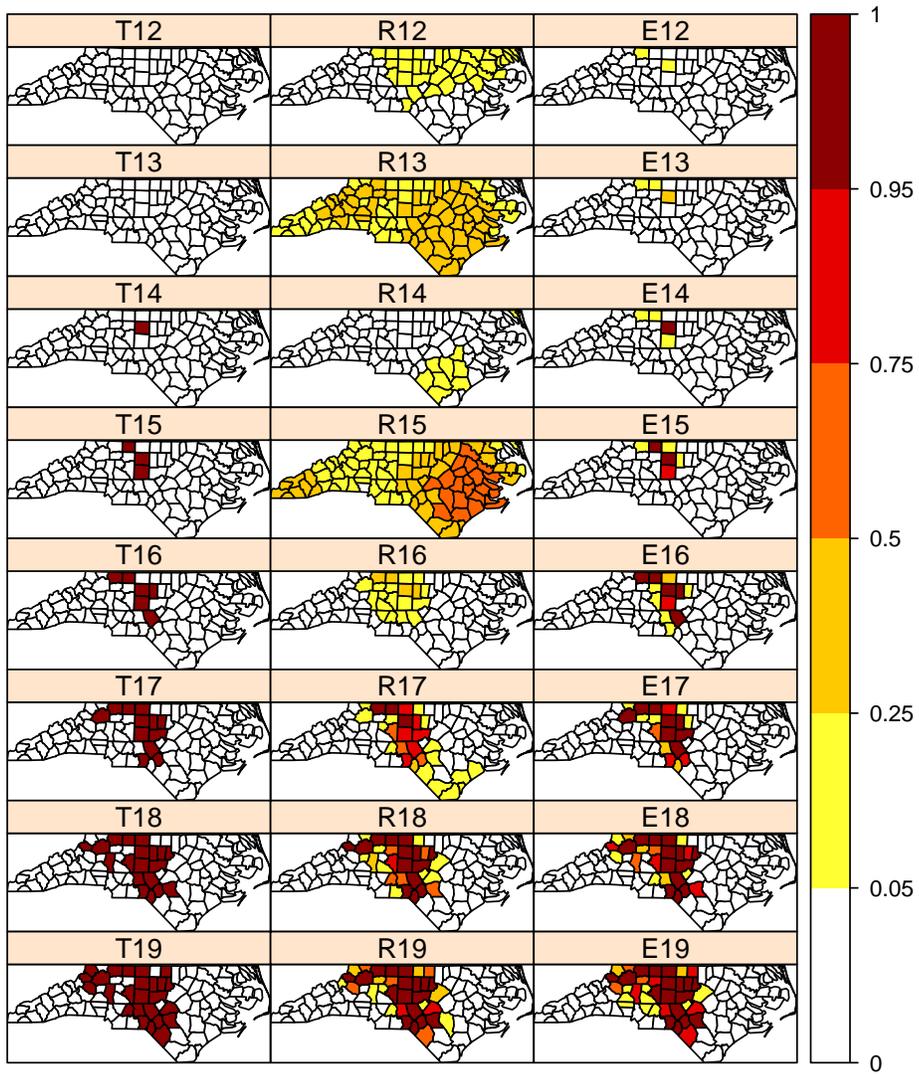
Figure 4: True vs Real time vs. *ex post facto* probabilities

a ROC-type (Receiver Operating Characteristic) curve for nine different thresholds ranging from 0.1 to 0.9. The x-axis stands for average maximum days delay for a given threshold over the 100 simulated datasets. The y-axis represents average false alarms for the real-time estimator over all the counties and replicates. We can see that the improvement in false alarms is obtained by raising the threshold, whereas resulting in increasing the delay of detection. However, the optimal threshold should depend on different focuses. The choice of a low threshold reflects an emphasis on timeliness, whereas choosing a high value stresses on low false alarm rate. This type of graph is valuable to public health authorities, as it enables decision makers to evaluate the situation and determine acceptable tradeoffs.
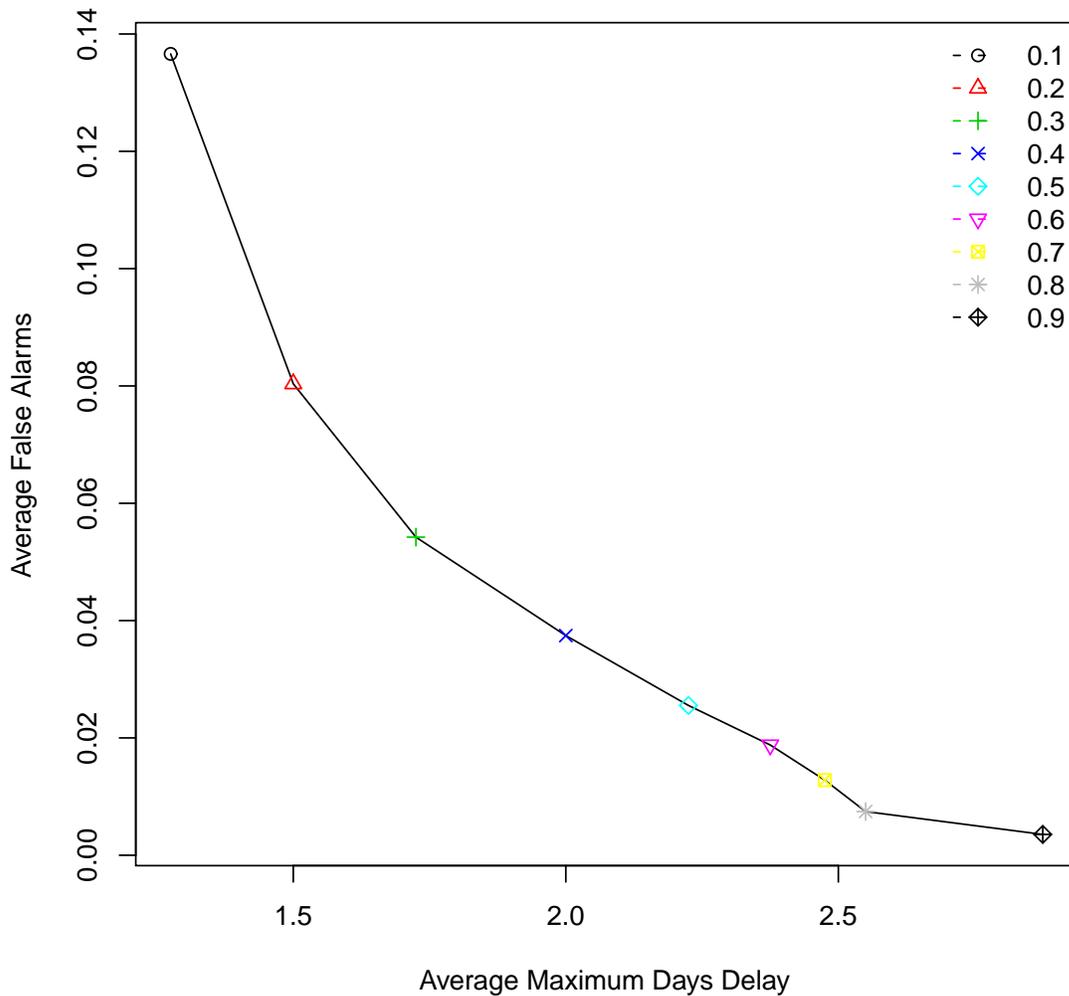
Figure 5: Average false alarms vs Average maximum number of days delayed

5 Discussion

In this paper, we have presented a new methodology that adapts the existing Gaussian Markov random field class of models to deal with spatio-temporal surveillance data. When the data are mainly spatial and coarsely discretized in time, simple models such as the CAR model will continue to be valuable for descriptive analysis. However, when data have a fine resolution in both the spatial and temporal dimensions, our model, which explicitly incorporates the directional nature of time by conditioning future events on past outcomes, is likely to be more insightful.

Our model for the latent stochastic process $\xi_i(t)$ is, essentially, a CAR model incorporating a sensible and computationally convenient Markov dependence structure in time. By contrast, most previous work only considers either space or time individually, which limits the model's ability to borrow strength from the other dimension. Inference for our relatively complex but realistic model requires computationally intensive Monte Carlo methods.

A natural extension of the model is to deal with multivariate counts. Then we can model the spatio-temporal random effect $\boldsymbol{\xi}(t)$ by a time-dependent variant MCAR model as in Banerjee, Carlin and Gelfand (2004), which assumes that conditional on the previous time period, $\boldsymbol{\xi}(t)$ at the current time follows a multivariate conditional autoregressive model. Another interesting extension of our proposed approach would be to include time-dependent predictors, $\boldsymbol{X}(t) = (x_1(t), \ldots, x_p(t))^T$, that potentially impact the joint distribution of $\boldsymbol{\lambda}$ and $\boldsymbol{Y}$. The model inference can be accomplished by a dynamic linear model framework (West and Harrison 1997).

Acknowledgments

References

[1] Banerjee, S., Carlin, B. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, CRC, New York.

[2] Banks, D., Datta, G., Karr, A., Lynch, J., Niemi, J. and Vera, F. (2010). "Bayesian CAR models for syndromic surveillance on multiple data streams: theory and practice," *Information Fusion*, In press.

[3] Barry, D. and Hartigan, J. (1993). "A Bayesian analysis for change point problems," *Journal of the American Statistical Association*, **88**, 309–319.

[4] Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, **36**, 192–225.

[5] Besag, J. (1975), "Statistical Analysis of Non-Lattice Data," *The Statistician*, **24**, 179–195.

[6] Besag, J. (1986), "On the Statistical Analysis of Dirty Pictures" (with discussion), *Journal of the Royal Statistical Society, Series B*, **48**, 259–302.

[7] Besag, J., York, J. C. and Mollie, A. (1991). "Bayesian image restoration, with two applications in spatial statistics (with discussion)," *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

[8] Besag, J. and Kooperberg, C. (1995). "On conditional and intrinsic autoregressions," *Biometrika*, **82(4)**, 733–746.

[9] Carlin, B., Gelfand, A. and Smith, A. F. M., (1992). "Hierarchical Bayesian analysis of change point problems," *Applied Statistics*, **41**, 389–405.

[10] Chib, S. (1998). "Estimation and comparison of multiple change point models," *Journal of Econometrics*, **86(2)**, 221–241.

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

[12] Cowling, B. J., Wong, I. O. L., Ho, L.-M., Riley, S., and Leung, G. M. (2006). "Methods for monitoring influenza surveillance data," *International Journal of Epidemiology*, **35**, 1314–1321.

[13] Fricker, R. D., Hegler, B. L., and Dunfee, D. A. (2008). "Comparing syndromic surveillance detection methods: EARS versus a CUSUM-based methology," *Statistics in Medicine*, **27**, 3407–3429.

[14] Giron, F., Moreno, E. and Casella, G. (2007). "An Objective Bayesian Analysis of the change point Problem," *Bayesian Statistics*, **8**, 1-27.

[15] Green, P. J., and Richardson, S. (2002). "Hidden Markov Models and Disease Mapping," *Journal of the American Statistical Association*, **97**, 1055–1070.

[16] Haario, H., Saksman, E., and Tamminen, J. (2001). "An Adaptive Metropolis Algorithm," *Bernoulli*, **7**, 223–242.

[17] Knorr-Held, L. and Richardson, S. (2003). "A Hierarchical Model for Space-Time Surveillance Data on Meningococal Disease Incidence," *Journal of the Royal Statistical Society Series C*, **52**, 169–183.

[18] Kulldorff, M. (1997). "A Spatial Scan Statistic," *Communications in Statistics–Theory and Methods*, **26**, 1481–1496.

[19] Martínez-Beneito, M. A., Conesa, D., López-Quílez, A., and López-Maside, A. (2008). "Bayesian Markov switching models for the early detection of influenza epidemics," *Statistics in Medicine*, **27**, 4455–4468.

[20] McCulloch, R. and Tsay, R. (1993). "Bayesian Inference of Trend and Difference-Stationarity," *Econometric Theory*, **10**, 596–608.

[21] Ogden, R. T. and Lynch, J. (1999). "Bayesian analysis of change point models," *Bayesian inference in Wavelet-based models,* Müller, P. and Vidakovic, B. (eds.), 67–82, New York: Springer-Verlag.

[22] Page, E. S. (1954). "Continuous Inspection Schemes," *Biometrika*, **41**, 100–115.

[23] Rossi, G., Lampugnani, L., and Marchi, M. (1999). "An approximate cusum procedure for surveillance of health events," *Statistics in Medicine*, **18**, 2111–2122.

[24] Smith, A. F. M. (1975). "A Bayesian approach to inference about a change point in a sequence of random variables," *Biometrika*, **62**, 407-416.

[25] Spiegelhalter, D., Grigg, O., Kinsman, R., and Treasure, T. (2003). "Risk-adjusted sequential probability raio tests: applications to Bristol, Shipman and adult cardiac surgery," *International Journal for Quality in Health Care*, **15**, 7–13.

[26] Sun, D., Tsutakawa, R.K., Kim, H. and He, Z. (2000). "Spatio-temporal interaction with disease mapping," *Statistics in Medicine*, **19**, 2015–2035.

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

[28] West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.

[29] Zhou, H. and Lawson, A. B. (2008). "EWMA smoothing and Bayesian spatial modeling for health surveillance," *Statistics in Medicine*, **27**, 5907–5928.