A Latent Model To Detect Multiple Clusters of Varying Sizes

Minge Xie, Qiankun Sun and Joseph Naus

Department of Statistics

Rutgers, the State University of New Jersey

Piscataway, NJ 08854

Summary

This paper develops a latent model and likelihood based inference to detect temporal clustering of events. The model mimics typical processes generating the observed data. We apply model selection techniques to determine the number of clusters, and develop likelihood inference and a Monte-Carlo EM algorithm to estimate model parameters, detect clusters and identify cluster locations. Our method differs from the classical scan statistic in that we can simultaneously detect multiple clusters of varying sizes. We illustrate the methodology with two real data applications and evaluate its efficiency through simulation studies. For the typical data generating process, our methodology is more efficient than a competing procedure that relies on least squares.

Some Key Words: AIC and BIC criteria; Clustering events; EM algorithm; Latent model; Likelihood inference; MCMC algorithm; Scan Statistics; Temporal samples.

Running Title: Detecting Clusters of Varying Sizes

1. Introduction

This paper develops a general latent modeling framework and likelihood based inference tools for detecting clustering of events within temporal samples. In the past 50 years, researchers have been investigating different types of clustering of events in time and space. Some applications look for an unusually large number of events in small clusters, or some patterns suggesting clumping over the entire study period or area. Other applications are concerned with unusually large clusters within a small region of time, space, or location in a sequence. In some cases, focus is on a specific region, for example a region with heavy pollution. In other cases, researchers scan the entire study area and seek to locate regions with unusually high likelihood of clustering. Practical examples cover a wide range of fields over various disciplines. In epidemiological studies when the "etiology of diseases" has not been well established, it is often required to analyze the data to obtain evidence of temporal clusters (Molinari, Bonaldi and Daures, 2001). In surveillance for biological terrorism, it is essential to provide early warnings of terrorist attacks. "Syndromic surveillance" for biological terrorism requires statistical methods of detecting "relatively abrupt increase in incidence" (Wallenstein and Naus, 2004). In environmental studies, people living near a factory generating pollution, may have an increased chance of certain diseases, and it is of interest to detect and monitor such clusters (Diggle, Rowlingson and Su, 2005). In biological studies of DNA sequencing, the detection of unusual clusters of specific patterns can be used to allocate lab resources and help find "biologically important origins of diseases" (Leung et al, 2005).

A traditional statistical method to detect a cluster of events is via *Scan Statistics*; see, e.g., Glaz, Naus, and Wallenstein (2001) and Fu and Lou (2003). The most commonly used scan statistic is the maximum number of cases in a fixed size moving-window that scans through the study area. The test based on this scan statistic has been shown to be a generalized likelihood ratio test for a uniform null against a pulse alternative. A related scan statistic is the diameter of the smallest window that contains a fixed number of cases. Other scan statistics and related likelihood based tests for localized temporal or spatial clustering are developed, that use a range

of fixed window sizes or a range of fixed number of cases. See, Kulldorff and Nagarwalla (1995), Naus and Wallenstein (2004), among others. One can also test for several unusually large clusters using asymptotic and approximate results; See, Dembo and Karlin (1992) and Su, Wallenstein and Bishop (2001). Gangnon and Clayton (2004) develop "a weighted average likelihood ratio scan statistic" and a"penalized scan statistic", which can be viewed as generalized scan statistics approaches.

Scan statistics procedures have been very successful in detecting a single significant cluster, and they also have had some success in detecting multiple clusters of fixed sizes. But there are some technical difficulties to detecting multiple clusters of varying sizes. In recent years, there have been several attempts to overcome this difficulty. The best approach so far is by Molinari et al (2001), who use a stepwise regression model together with model selection procedures to locate and determine the number of clustering regions in temporal data. For a given number of clusters, the locations of the clusters are determined by a least square method where the response variable is the set of inter-arrival times (gaps) between events. To make inference, they rely on bootstrap methodology and the least square formulation. Because the responses used in their model are usually non-normally distributed, the least square method may not be efficient. Also, the bootstrap simulation could be computationally expensive. To overcome the difficulty of using bootstrap simulation, Demattei and Molinari (2006) add a new testing method to Molinari et al's (2001) stepwise regression method. The new testing method utilizes Bernstein's inequality, resulting in a conservative test. This stepwise regression method is extended by Demattei, Molinari and Daures (2007) to detect arbitrarily shaped multiple clusters in spatial data.

Diggle et al (2005) develop a two-level hierarchical model based on a latent stochastic process to detect localized regions of unusually high likelihood of event (relative to background) in temporal-spatial data. They indicate how a third level of hierarchy, a model for the prior distribution of the parameters of the latent stochastic process, can be incorporated for a Bayesian analysis. The model in Diggle et al (2005) approach is useful in many applications. It is very

similar to the so-called "disease mapping" approach (e.g., Clayton and Kaldor, 1987, Besag, York and Mollie, 1991, Waller et al, 1997, and Gangnon and Clayton, 2000). Note, however, that the goal of Diggle et al. (2005) is to describe intensity functions instead of directly detecting and making inference on clusters.

Neill, Moore and Cooper (2006) propose a Bayesian scan statistic for detecting spatial clusters, and compare their method to a frequentist spatial scan statistic method (Kulldorff and Nagarwalla 1995). By using conjugate priors, they can obtain closed form solutions for likelihood functions, which leads to a "much faster" algorithm. As in Kulldorff and Nagarwalla (1995) and many other papers, the potential spatial clusters are limited to a finite set of specific choices. Other Bayesian methods of detection of clusters include Lawson (1995), Gangnon and Clayton (2000, 2003), Knorr-Held and RaBer (2000), Denison and Holmes (2001). Although the Bayesian methods may allow us to incorporate prior information from experts in the domain, the choice of priors is always a "challenging task" (Neill et al, 2006).

Our method deals with the same problems as Molinari et al (2001). The approach that we use is very different. By mimicking the processes and mechanisms that generate clusters, we develop a latent model that allows us to use the standard likelihood inference to detect multiple clusters in a given time window. Unlike the scan statistics procedures, we focus on detecting multiple clusters of varying sizes. Our method can be used to simultaneously detect multiple clusters, as well as a significant single cluster. In our approach the likelihood function can be fully specified. We can answer a variety of inference questions related to our goal. Under the assumed model, the likelihood based approach is more efficient than the stepwise regression approach by Molinari et al (2001), which is illustrated in our simulation studies. Furthermore, our latent model approach is very flexible and it can incorporate several extensions (Sun, 2008).

The rest of the paper is arranged as follows. Section 2 proposes a general latent model for multiple clusters, that generates the observed data (observed time points of an event). Section 3 develops, for a given number of clusters, approaches based on likelihood inference to estimate

and test model parameters and detect clusters. Section 4 employs model selection approaches, both *A*kaike (AIC) and *B*ayesian (BIC) *I*nformation *C*riteria, to determine the number of clusters. Section 5 contains two real data analysis examples, including analysis of the Hospital Hemoptysis Admission data studied by Molinari et al (2001) and a set of brucellosis event data collected by the CDC (*C*enters for *D*isease *C*ontrol and Prevention). Section 6 provides a comprehensive simulation study to further illustrate and evaluate the proposed methodology. Section 7 provides some additional comments and discussions.

2. A latent multiple cluster model

Suppose in a given time window, say $(0, T)$, there are k clusters, where k is a fixed integer. Here, clusters are defined as the time intervals within which an event of interest is much more (or less) likely to happen (per unit of time) than outside these time intervals. A temporal (latent) clustering model is specified as in Figure 1. Starting from time 0, we wait for b_1 units of time until the first cluster appears, and the first cluster lasts c_1 units of time. After the first cluster, we wait for b_2 units of time until the second cluster appears, and the second cluster lasts c_2 units of time, and so on until the kth cluster appears that lasts c_k units of time. After the kth cluster, b_{k+1} is the waiting period till the next cluster, which occurs after the endpoint T. From now on, we will use this clustering model to illustrate the development of our methodology. Other potential latent clustering models are discussed in Section 7.

[Insert Figure 1 here]

To complete the model specification, we further assume that the waiting time periods $b_1, b_2, \ldots, b_{k+1}$ and the cluster interval lengths $c_1, c_2, \ldots, c_k$ are random variables. We assume that $b_1, b_2, \ldots, b_{k+1}$ are independent samples from a distribution with a density function $\psi_b(t) = \psi_b(t; \lambda_b)$ and $c_1, c_2, \ldots, c_k$ are independent samples from a distribution with a density function $\psi_c(t) = \psi_c(t; \lambda_c)$. Here, λ_b and λ_c are unknown parameters, ψ_b and ψ_c are known density functions that may or may not be from the same family of distributions. One simple example is that both $\psi_b(t)$ and $\psi_c(t)$ are exponential densities but with means equal to $1/\lambda_b$ and $1/\lambda_c$, respectively.

Write $\mathbf{b} = (b_1, b_2, \ldots, b_{k+1})'$ and $\mathbf{c} = (c_1, c_2, \ldots, c_k)'$. It can be seen from Figure 1 that $I_j = [\sum_{l=1}^{j-1}(b_l + c_l) + b_j, \sum_{l=1}^{j}(b_l + c_l)]$ is the interval of the jth cluster, $j = 2, \ldots, k$, and $I_1 = (b_1, b_1 + c_1)$. For convenience, we introduce a random number δ such that $\{\delta = k\}$ is the event that exactly k clusters occur in time window $(0, T)$. Clearly, event $\{\delta = k\}$ is equivalent to event $\{\sum_{j=1}^{k}(b_j + c_j) + b_{k+1} \geq T$ and $\sum_{j=1}^{k}(b_j + c_j) \leq T\}$.

The latent variables $\mathbf{b}$ and $\mathbf{c}$ are not observed. What we can observe in this model setting are only the time points $y_1, y_2, \ldots, y_n$ when an event of interest occurs. We assume that the observations $y_1, y_2, \ldots, y_n$ are independent samples from the following step uniform function,

$$f_\theta(y|\mathbf{b}, \mathbf{c}, k) = \begin{cases} \frac{\alpha_1}{T+\sum_{j=1}^{k}(\alpha_j - 1)c_j}, & \text{if } y \in I_1; \\ \quad \ldots\ldots \\ \frac{\alpha_k}{T+\sum_{j=1}^{k}(\alpha_j - 1)c_j}, & \text{if } y \in I_k; \\ \frac{1}{T+\sum_{j=1}^{k}(\alpha_j - 1)c_j}, & \text{if } y \notin \cup_{j=1}^{k} I_j, \end{cases} \tag{1}$$

where $\boldsymbol{\theta} = (\alpha^T, \lambda^T)'$ is the collection of all parameters, including unknown parameters $\alpha = (\alpha_1, \ldots, \alpha_k)'$ and the parameters $\lambda = (\lambda_b, \lambda_c)'$ that are associated with random variables b_i's and c_i's. When $k = 1$, the step uniform density function (1) becomes the single step uniform density function used for the single cluster case; See, e.g., Chapter 14 of Glaz, Naus, and Wallenstein (2001). The parameters $\alpha_j > 0$ for each $j = 1, 2, ..., k$; they may or may not be the same across the k clusters. Under the density assumption (1), the event is α_j times more likely to happen (per unit of time) inside the jth cluster than that outside the clusters. The case with $\alpha_j > 1$ corresponds to a more dense cluster of more events; the case with $\alpha < 1$ corresponds to a more sparse cluster of less events; and the case with $\alpha = 1$ corresponds to no cluster. To see whether there are any significant clusters in the data, we can test a hypothesis H_0: $\alpha_1 = \alpha_2 = ... = \alpha_k = 1$ versus H_1: at least one $\alpha_j \neq 1$.

The proposed distribution assumption can be alternatively expressed in terms of Poisson models, similar to those used in Gangnon and Clayton (2004) and others. We use the current formulation of a step uniform function in order to highlight the interpretation of the parameters

α_j's. The proposed model is also closely related to a Bayesian model. In particular, if we further assume that α_j's are random and place priors on the parameters, the proposed model corresponds to a Bayesian hierarchical model. We use the frequentist formulation, since it allows us to utilize the fully developed likelihood inferences and avoid choosing priors. Although we illustrate our latent model for temporal data, the model developed also covers other types of data, for example, with patterns or events in a sequence such as the DNA data studied by Leung et al (2005).

We also consider a slight generalization of model (1), that includes a known background function $W(t)$. As mentioned by Molinari et al (2001) and Wallenstein and Naus (2004), the background value, such as seasonal patterns or population sizes, may not be the same across the time window $(0, T)$. The known background function $W(t)$, usually assessed from separate sources, can be easily incorporated into our model. In this case, we replace (1) by

$$f_\theta(y|\mathbf{b}, \mathbf{c}, k) = \begin{cases} \frac{\alpha_1 W(y)}{\tilde{T}+\sum_{j=1}^{k}(\alpha_j-1)\tilde{c}_j}, & \text{if } y \in I_1 \\ \quad \cdots\cdots & \\ \frac{\alpha_k W(y)}{\tilde{T}+\sum_{j=1}^{k}(\alpha_j-1)\tilde{c}_j}, & \text{if } y \in I_k \\ \frac{W(y)}{\tilde{T}+\sum_{j=1}^{k}(\alpha_j-1)\tilde{c}_j}, & \text{if } y \notin \cup_{j=1}^{k} I_j \end{cases} \tag{2}$$

where $\tilde{T} = \int_0^T W(t)dt$ and $\tilde{c}_j = \int_{I_j} W(t)dt$, for $j = 1, 2, \ldots, k$. Fitting model (2) is exactly the same as fitting model (1), except that T and c_j need to be replaced by $\tilde{T}$ and $\tilde{c}_j$. To simplify our presentation, we focus our developments in the next few sections on model (1).

3. Model inference for a given number of clusters

The number of clusters k needs to be bounded away from the number of observations n, so that there are no overfitting problems such as the case of a cluster consisting of a single event. We apply model selection techniques to determine k; for a fixed k, we develop a likelihood approach to estimate model parameters and cluster locations and make statistical inference. We assume that k is known in this section, and discuss how to determine k in Section 4.

3.1. Likelihood function of observed data when $\delta = k$.

In the latent model illustrated in Figure 1, the probability of k clusters existing in the time window $(0, T)$ can be computed by

$$P_\lambda(\delta = k) = P_\lambda\Big\{ \sum_{j=1}^{k}(b_j + c_j) + b_{k+1} \geq T \text{ and } \sum_{j=1}^{k}(b_j + c_j) \leq T \Big\} = \int_0^T \int_{T-s}^{\infty} \psi_b(t)\psi_{bc}^{[k]}(s)dtds.$$

where $\psi_{bc}^{[k]}(s)$ is the density function of $\sum_{j=1}^{k}(b_j + c_j)$. Conditional on $\delta = k$, the joint conditional density function of $(\mathbf{b}, \mathbf{c})$ is

$$f_\theta(\mathbf{b}, \mathbf{c}|k) = \prod_{j=1}^{k}\{\psi_b(b_j)\psi_c(c_j)\}\psi_b(b_{k+1})/\{P_\lambda(\delta = k)\}.$$

From model (1), the conditional joint density function of the sample observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, conditional on $\mathbf{b}$, $\mathbf{c}$ and $\delta = k$, is

$$f_\theta(\mathbf{y}|\mathbf{b}, \mathbf{c}, k) = \prod_{i=1}^{n} f(y_i|\mathbf{b}, \mathbf{c}, k) = e^{\sum_{j=1}^{k}(\log \alpha_j)Z_j - n\log\{T + \sum_{j=1}^{k}(\alpha_j - 1)c_j\}},$$

where $Z_j = Z_j(\mathbf{y}, \mathbf{b}, \mathbf{c}) = \sum_{i=1}^{n} 1_{(y_i \in I_j)}$ is the number of events that appear within the jth cluster interval. Here, $1_{(\cdot)}$ is the indicator function. Thus, the joint density function of $\mathbf{y}$ and $\delta = k$, is

$$f_\theta(\mathbf{y}, k) = \int\int f_\theta(\mathbf{y}, \mathbf{b}, \mathbf{c}|k)P_\lambda(\delta = k)d\mathbf{b}d\mathbf{c} = \int\int f_\theta(\mathbf{y}|\mathbf{b}, \mathbf{c}, k)f_\theta(\mathbf{b}, \mathbf{c}|k)P_\lambda(\delta = k)d\mathbf{b}d\mathbf{c}$$

and the log-likelihood function of observing $\mathbf{y}$ and $\delta = k$ is

$$\ell_k(\theta|\mathbf{y}) = \log\{f_\theta(\mathbf{y}, k)\} = \log\Big\{ \int\int f_\theta(\mathbf{y}|\mathbf{b}, \mathbf{c}, k)f_\theta(\mathbf{b}, \mathbf{c}|k)P_\lambda(\delta = k)d\mathbf{b}d\mathbf{c}\Big\}. \tag{3}$$

Since (3) involves multiple integrations, it is complicated to directly compute the log-likelihood function $\ell_k(\boldsymbol{\theta}\,|\mathbf{y})$ and its first and second derivatives. As a result, it is hard to obtain the maximum likelihood estimates by directly maximizing the likelihood function. We instead develop next a Monte-Carlo EM algorithm (See, e.g., Tanner, Section 4.5) to estimate the model parameters.

3.2. A Monte-Carlo EM algorithm for model estimation.

Note that the joint density function of $(\mathbf{y}, \mathbf{b}, \mathbf{c}, \delta = k)$ is explicit,

$$\begin{aligned} f_\theta(\mathbf{y}, \mathbf{b}, \mathbf{c}, k) &= f_\theta(\mathbf{y}|\mathbf{b}, \mathbf{c}, k)f_\theta(\mathbf{b}, \mathbf{c}|k)P_\lambda(\{\delta = k\}) \\ &= e^{\sum_{j=1}^{k}(\log \alpha_j)Z_j - n\log\{T + \sum_{j=1}^{k}(\alpha_j - 1)c_j\}} \prod_{j=1}^{k}\{\psi_b(b_j)\psi_c(c_j)\}\psi_b(b_{k+1}). \end{aligned}$$

We treat $(\mathbf{y}, \mathbf{b}, \mathbf{c}, \delta = k)$ as the complete responses and $(\mathbf{y}, \delta = k)$ as the observed responses, and develop an EM algorithm as follows.

Step 0. Select a set of starting parameter values $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)\prime}, \lambda^{(0)\prime})'$.

Step 1 (E-step). For $s = 0, 1, 2, \ldots$, calculate the conditional expectation of the complete log-likelihood function, given the observed and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(s)}$: $Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(s)}) = Q_1(\alpha|\, \boldsymbol{\theta}^{(s)}) + Q_2(\lambda|\, \boldsymbol{\theta}^{(s)})$ where $Q_1(\alpha|\, \boldsymbol{\theta}^{(s)}) = \sum_{j=1}^{k} \mathrm{E}(Z_j|\mathbf{y}, k, \boldsymbol{\theta}^{(s)}) \log \alpha_j - n\mathrm{E}[\log\{T + \sum_{j=1}^{k}(\alpha_j - 1)c_j\}|\mathbf{y}, k, \boldsymbol{\theta}^{(s)}]$ and $Q_2(\lambda|\, \boldsymbol{\theta}^{(s)}) = \sum_{j=1}^{k+1} \mathrm{E}\{\log \psi_b(b_j)|\mathbf{y}, k, \boldsymbol{\theta}^{(s)}\} + \sum_{j=1}^{k} \mathrm{E}\{\log \psi_c(c_j)|\mathbf{y}, k, \boldsymbol{\theta}^{(s)}\}$.

Step 2 (M-step). For $s = 0, 1, 2, \ldots$, update the parameter estimates: $\boldsymbol{\theta}^{(s+1)} = (\alpha^{(s+1)}, \lambda^{(s+1)})'$, by maximizing the $Q_1(\alpha|\, \boldsymbol{\theta}^{(s)})$ and $Q_2(\lambda|\, \boldsymbol{\theta}^{(s)})$ functions: $\alpha^{(s+1)} = \mathrm{argmax}\, Q_1(\alpha|\, \boldsymbol{\theta}^{(s)})$ and $\lambda^{(s+1)} = \mathrm{argmax}\, Q_2(\lambda|\, \boldsymbol{\theta}^{(s)})$.

Step 3. Repeat Steps 2 and 3 until $\| \boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)} \|$ is very small.

In the case with ψ_b and ψ_c being density functions of exponential distributions $\mathrm{Exp}(\lambda_b)$ and $\mathrm{Exp}(\lambda_c)$, the updating formula of $\lambda^{(s+1)}$ in Step 2 is simply $\lambda_b^{(s+1)} = (k+1)/\sum_{j=1}^{k+1} \mathrm{E}(b_j|\mathbf{y}, k, \boldsymbol{\theta}^{(s)})$ and $\lambda_c^{(s+1)} = k/\sum_{j=1}^{k} \mathrm{E}(c_j|\mathbf{y}, k, \boldsymbol{\theta}^{(s)})$.

The conditional expectations in Step 1 do not usually have explicit forms, but they can be numerically computed using a Gibbs sampling approach. Suppose $\mathbf{b}^* = (b_1^*, b_2^*, \ldots, b_{k+1}^*)'$ and $\mathbf{c}^* = (c_1^*, c_2^*, \ldots, c_k^*)'$ are a set of Gibbs samples from $f(\mathbf{b}, \mathbf{c}|\mathbf{y}, k, \boldsymbol{\theta}^{(s)})$, and we have M sets of such Gibbs samples; see Appendix I in Supplementary Materials for a Gibbs sampling algorithm to generate $\mathbf{b}^*$ and $\mathbf{c}^*$. When M is large, the four conditional expectations in Step 1 of the EM algorithm can be evaluated by $\sum_* Z_j^*/M$, $\sum_* \log\{T + \sum_{j=1}^{k}(\alpha_j - 1)c_j^*\}/M$, $\sum_* \log\{\psi(b_j^*)\}/M$, and $\sum_* \log\{\psi(c_j^*)\}/M$, respectively, where $\sum_*$ is the summation over M sets of Gibbs samples $\mathbf{b}^*$ and $\mathbf{c}^*$, and Z_j^* is the total number of events, but computed with b_j and c_j values replaced by their corresponding Gibbs sample values b_j^* and c_j^* in each Gibbs sample set.

The above Monte-Carlo EM algorithm does not provide the variance-covariance calculation for the parameter estimators. To obtain an estimate of the variance-covariance matrix, we

calculate the observed information matrix using the missing information principle and Louis's Method (see, e.g. Tanner, 1993, Sections 4.4.2 - 4.4.3). In particular, the observed information matrix is $H_n = -\{\frac{\partial^2}{\partial\theta^2}\ell_k(\boldsymbol{\theta}\,|\mathbf{y})\} = -\mathrm{E}\{\frac{\partial^2}{\partial\theta^2}\ell_k(\boldsymbol{\theta}\,|\mathbf{b},\mathbf{c},\mathbf{y})|\mathbf{y},\delta=k\} - \mathrm{Var}\{\frac{\partial}{\partial\theta}\ell_k(\boldsymbol{\theta}\,|\mathbf{b},\mathbf{c},\mathbf{y})|\mathbf{y},\delta=k\}$, where $\ell_k(\boldsymbol{\theta}\,|\mathbf{b},\mathbf{c},\mathbf{y}) = \log\{f_\theta(\mathbf{y},\mathbf{b},\mathbf{c},k)\}$. It can be numerically estimated by

$$\hat{H}_n = -\frac{1}{M}\sum_{*}\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_k(\boldsymbol{\theta}\,|\mathbf{b}^*,\mathbf{c}^*,\mathbf{y}) - \Big[\frac{1}{M}\sum_{*}\Big\{\frac{\partial}{\partial\boldsymbol{\theta}}\ell_k(\boldsymbol{\theta}\,|\mathbf{b}^*,\mathbf{c}^*,\mathbf{y})\Big\}^2 - \Big\{\frac{1}{M}\sum_{*}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_k(\boldsymbol{\theta}\,|\mathbf{b}^*,\mathbf{c}^*,\mathbf{y})\Big\}^2\Big],$$

where the summations are over the M sets of Gibbs samples of $\mathbf{b}^*$ and $\mathbf{c}^*$ in the last iteration of the EM algorithm (See, e.g., Tanner, 1993, section 4.4.4.).

3.3. Estimation of cluster locations.

We often like to know where the clusters are located. Based on the model specification, the lower and upper bounds of the jth cluster interval I_j are respectively $\mathrm{L}_j = \sum_{l=1}^{j-1}(b_l + c_l) + b_j$ and $\mathrm{U}_j = \sum_{l=1}^{j}(b_l + c_l)$. They are random quantities that can not be estimated by maximizing the observed likelihood function (3). Fortunately, based on the M sets of the Gibbs samples in the last iteration of the EM algorithm, we are able to obtain M copies of the pair $(\mathrm{L}_j^*, \mathrm{U}_j^*) = (\sum_{l=1}^{j-1}(b_l^* + c_l^*) + b_j^*, \sum_{l=1}^{j}(b_l^* + c_l^*))$. These $(\mathrm{L}_j^*, \mathrm{U}_j^*)$ can be treated as samples from the conditional joint distribution ("posterior distribution") of $(\mathrm{L}_j, \mathrm{U}_j)$, given $\mathbf{y}$ and k. Thus, following Bayesian point estimation techniques, we use the mode of the empirical conditional joint distribution of $(\mathrm{L}_j, \mathrm{U}_j)$, obtained from the M pairs of $(\mathrm{L}_j^*, \mathrm{U}_j^*)$, to estimate the random bounds L_j and U_j. See Fraiman and Meloche (1999) for a kernel based estimation approach to estimate the mode of a bivariate distribution from its M copies of samples.

3.4. Likelihood inference and simulation based tests related to α's.

For a fixed k, we can obtain k clusters from the aforementioned estimation algorithm. There is no guarantee that the k clusters are statistically significant. To test whether an identified cluster is significant or not, we use likelihood based inference.

Let us first consider testing a single (say the jth) cluster and see whether it is significant or not, i.e., $H_0 : \alpha_j = 1$ versus $H_1 : \alpha_j \neq 1$. Let $\hat{\alpha}_j$ be the estimator of the parameter α_j. From the

observed information matrix $\hat{H}_n$, we can get an estimator of the standard error of $\hat{\alpha}_j$, say $\hat{se}(\hat{\alpha}_j)$. Thus, a Wald-type t statistic is $t = \hat{\alpha}_j/\hat{se}(\hat{\alpha}_j)$. When n is large, the statistic t is asymptotically normally distributed based on likelihood inference and we can use a two-sided z test to test for whether $\alpha_j = 1$ or not.

Another testing problem of interest is to see whether there are any significant clusters among the k clusters, i.e., $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 1$ versus H_1 : at least one $\alpha_j \neq 1$. The likelihood ratio test statistic is

$$\begin{aligned} R &= \log\left\{\frac{\max_{H_1\cup H_0} f_\theta(\mathbf{y},k)}{\max_{H_0} f_\theta(\mathbf{y},k)}\right\} = \ell_k(\hat{\boldsymbol{\theta}}|\mathbf{y})|_{\theta=\hat{\theta}} + n\log(T) - \max_{\lambda}\log P_\lambda(\delta=k) \\ &= \log\int\int f_{\hat{\theta}}(\mathbf{y}|\mathbf{b},\mathbf{c},k) f_{\hat{\theta}}(\mathbf{b},\mathbf{c}|k) d\mathbf{b} d\mathbf{c} + \log P_{\hat{\lambda}}(\delta=k) + n\log(T) - \max_{\lambda}\log P_\lambda(\delta=k), \end{aligned}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\alpha}', \hat{\lambda}')'$ are the estimates of the parameters obtained from the aforementioned EM algorithm under H_1.

Suppose we have M sets of random samples $\mathbf{b}^{**} = (b_1^{**}, \ldots, b_{k+1}^{**})$ and $\mathbf{c}^{**} = (c_1^{**}, \ldots, c_k^{**})$ from $f_\theta(\mathbf{b},\mathbf{c}|k)$ when $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$; see Appendix I in Supplementary Materials for a Gibbs sampling algorithm to simulate such $\mathbf{b}^{**}$ and $\mathbf{c}^{**}$. By Monte-Carlo approximation, the test statistic R can be approximated by

$$R^{**} = \log\{\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{b}^{**},\mathbf{c}^{**},k)\} + \log P_{\hat{\lambda}}(\delta=k) + n\log(T) - \max_{\lambda}\log P_\lambda(\delta=k),$$

where $\sum_{**}$ is the summation over the M sets of $\mathbf{b}^{**}$ and $\mathbf{c}^{**}$ samples. Based on likelihood inference, $2R$ is asymptotically χ^2 distributed with k degrees of freedom. By comparing $2R^{**}$ with the χ^2_k distribution, we perform a formal test for $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 1$ versus H_1 : at least one $\alpha_j \neq 1$.

The above test relies on large sample theory and requires a large n. Alternatively, we consider a simulation based Monte-Carlo testing approach that is computationally intensive. We simulate L sets of data of samples of size n from the model under the null hypothesis and compute for each set the test statistic $2R^{**}$, denoted by $2\tilde{R}^{**}$. When L is large, the empirical distribution of these

$2\tilde{R}^{**}$ values provides a good approximation to the theoretical distribution of the test statistic $2R^{**}$ under the null hypothesis. This is utilized to perform the simulation based Monte-Carlo test.

4. Determination of the unknown number of clusters

In this section, we employ model selection approaches, both Akaike (AIC) and Bayesian (BIC) information criteria, to determine the number of clusters k from the observed data. In our context, their expressions are $\text{AIC}(k) = -2\log\{\int\int f_\theta(\mathbf{y}|\mathbf{b},\mathbf{c},k)f_\theta(\mathbf{b},\mathbf{c}|k)d\mathbf{b}d\mathbf{c}\} - 2\log P_\lambda(\delta = k) + 2k$ and $\text{BIC}(k) = -2\log\{\int\int f_\theta(\mathbf{y}|\mathbf{b},\mathbf{c},k)f_\theta(\mathbf{b},\mathbf{c}|k)d\mathbf{b}d\mathbf{c}\} - 2\log P_\lambda(\delta = k) + k\log(n)$, respectively. Often the number of observed time points $n > e^2 = 7.389$, thus the BIC criterion places more penalty against a large k than the AIC criterion.

To compute the criteria, the unknown parameters $\boldsymbol{\theta}$ are replaced by their estimates $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(k)$ that are obtained from the Monte-Carlo EM algorithm proposed in Section 3.2 (set the number of clusters to be k). Furthermore, the formulas of the criteria involve integrations that do not have explicit forms. So, we numerically evaluate their values. As in Section 3.4, we know how to simulate $\mathbf{b}^{**} = (b_1^{**},\ldots,b_{k+1}^{**})$ and $\mathbf{c}^{**} = (c_1^{**},\ldots,c_k^{**})$ from $f_\theta(\mathbf{b},\mathbf{c}|k)$ when $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(k)$. By Monte-Carlo approximation, the $\text{AIC}(k)$ criterion can be computed by

$$\widehat{\text{AIC}}(k) = -2\log\{\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{b}^{**},\mathbf{c}^{**},k)\} - 2\log P_{\hat{\lambda}}(\delta = k) + 2k,$$

and the $\text{BIC}(k)$ criterion can be computed by

$$\widehat{\text{BIC}}(k) = -2\log\{\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{b}^{**},\mathbf{c}^{**},k)\} - 2\log P_{\hat{\lambda}}(\delta = k) + k\log(n),$$

where $\sum_{**}$ is the summation over M sets of repeatedly simulated $(\mathbf{b}^{**},\mathbf{c}^{**})$'s from $f_{\hat{\theta}(k)}(\mathbf{b},\mathbf{c}|k)$. The k to be chosen is the one with the smallest corresponding $\widehat{\text{AIC}}(k)$ or $\widehat{\text{BIC}}(k)$ value.

Based on the developments in Sections 3 and 4, a practical approach for detecting clusters emerges. Denote by $\mathcal{K}$ a pre-selected set of k's, which we would like to be small for computing purposes but large enough to cover all potential choices of the correct number of clusters. For

each $k \in \mathcal{K}$, apply the Monte-Carlo EM algorithm in Section 3.2 to get the parameter estimates and use either the AIC or BIC rule to determine the number of clusters k. For the chosen k, use the results in Sections 3.3-3.4 to detect and determine the location of the clusters.

5. Real data examples

We analyze two real data sets in this section. The first is the hospital hemoptysis admission data studied by Molinari et al (2001). The second is the brucellosis data collected by the CDC during 1997-2004. Both data sets are attached in Appendix II in Supplementary Materials.

5.1. Hospital hemoptysis admission data set.

The first data set consists of 62 spontaneous hemoptysis admissions (pulmonary disease) at Nice (a southern French city) hospital from January 1 to December 31, 1995. Since Nice is a tourist city located on the Mediterranean coast, many tourists ($\sim 15\%$ of the local population) each summer increase the population at risk. Molinari et al (2001) suggest using the following function to adjust for the population at risk $R(t) = 1 + \frac{72t}{10{,}000 \times 365} + \frac{55{,}000}{355{,}000} \times \mathbf{1}_{[182,244]}(t), t \in [1, 365]$.

We reanalyze the data using our method for k's in the set $\mathcal{K} = \{1, 2, 3, 4\}$ both with and without incorporating the background function $R(t)$. Both the AIC and BIC criteria select $k = 1$ regardless of whether the background function is used or not. Part I of Table 1 summarizes the results of estimation, testing and cluster detection in the cases of using $k = 1$ and $k = 2$.

[Insert Table 1 here]

Included in Table 1 are also the results from the stepwise regression (SR) approach by Molinari et al (2001) and a varying size window scan statistic approach using the free SaTScan software (www.satscan.org). The results of the stepwise regression approach are reported in Molinari et al (2001) with a constraint that a cluster should have a minimum of 6 events. The p-value marked with † is reported in a follow up paper by Demattei and Molinari (2006), in which they state that the bootstrap based inference by Molinari et al (2001) is "not reliable".

Based on Table 1, all three methods point to one potential non-significant cluster. Our estimated cluster interval is from Day 58 to Day 108 (February 27-April 18). The interval covers,

but is a little longer than, the estimated cluster interval of Day 58 to Day 87 (February 27-March 28) by the stepwise regression and the SaTScan methods.

5.2. CDC Brucellosis data.

Brucellosis (Malta fever) is an infectious disease transmitted from animals to human. It is caused by bacteria of the genus Brucella and is one "critical biologic agent reported to NNDSS (National Notifiable Disease Surveillance System)" (Chang et al 2003). The data considered here is weekly events across the US, collected every year by the CDC. We analyze the 2004 weekly data, using a background function estimated by the weekly number of brucellosis cases averaged over year 1997 to year 2003.

Since the data is provided as weekly counts, to avoid simultaneously occurring events, we uniformly disperse the multiple cases in a week and then analyze the transformed data. We analyze the data using our method for k's in the set $\mathcal{K} = \{1, 2, 3, 4\}$. When the background function is not used, the AIC criterion selects $k = 2$ and the BIC criterion selects $k = 1$. When the background function is used, both the AIC and BIC criterions select $k = 1$. Part II of Table 1 lists the results of model fitting, testing and cluster detection in the cases of $k = 1$ and $k = 2$ for the CDC Brucellosis data.

Included in Table 1, for comparison, are also the results from the varying size window scan statistic using the SaTScan software. Since the primary cluster is significant, we have also tried to use the SaTScan software to pick up the secondary cluster by replacing the event counts in the primary cluster (week 44 to week 46) with the average outside the primary cluster. However, it failed to find any additional significant cluster.

Based on these results, we can conclude that there exists one significant cluster from week 44 to week 46 (October 30-November 19) in the 2004 Brucellosis data.

6. Simulation studies

In this section, we perform simulation studies from two settings of pre-fixed clusters to evaluate the performance of the proposed estimation, testing and cluster detection methods. Without

loss of generality, all simulation studies are done within the time window $(0, 1)$.

Our first simulation setting fixes a single cluster at $(.258, .494)$, which is based on a realization of the proposed latent cluster model with $k = 1$. Choosing $\alpha = 3$, we simulate 300 sets of size $n = 100$ independent time points $y_1, y_2, \ldots, y_{100}$ according to model (1). The second simulation setting is for a multiple cluster case with $k = 3$. In particular, based on a realization of the latent cluster model with $k = 3$, we fix the cluster intervals at $(.092, .283)$, $(.490, .584)$ and $(.743, .891)$. We then choose $\alpha = (\alpha_1, \alpha_2, \alpha_3)' = (3, 3.25, 3.5)'$, and simulate 300 sets of size $n = 180$ independent time points $y_1, y_2, \ldots, y_{180}$ according to model (1). We assume that we can only observe the y values in our data analysis. We fit each of the 2×300 data sets first with k being their respective true number of clusters. Table 2 summarizes the results of parameter estimation, cluster detection and testing. For comparison purpose, we also include in Table 2 the corresponding results from the stepwise regression (SR) method proposed by Molinari et al (2001) and Demattei and Molinari (2006).

[Insert Table 2 here]

The first part of Table 2 provides the mean values of 300 estimates of the main parameters α (log-transformed) and their standard errors. It indicates that the Monte-Carlo EM algorithm has provided very reasonable estimates for the parameters α. The second part of Table 2 employs four empirical measures to assess the accuracy of the estimated cluster locations: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Here, sensitivity is the proportion of the event points (y's) inside the true clusters, that are inside the estimated clusters. Specificity is the proportion of the event points (y's) outside the true clusters, that are outside the estimated clusters. PPV is the proportion of the event points (y's) inside the estimated clusters, that are inside the true clusters. NPV is the proportion of the event points (y's) outside the estimated clusters, that are outside true clusters. The closer these measures are to one, the more accurate the cluster intervals. Reported in Table 2 are the means and standard deviations of the sensitivity, specificity, PPV and NPV values in the 300 repeated simulations for

the cluster location estimation method described in Section 3.3 as well as the stepwise regression method proposed by Molinari et al (2001) and Demattei and Molinari (2006). The results indicate both our and the stepwise regression methods can identify clusters reasonably, but all four measures in our methods are almost uniformly higher with smaller standard deviations.

The third part of Table 2 examines the powers and sizes of the level .05 tests outlined in Section 3.4. For the power calculation, we use the same 2×300 simulation data sets described before. To compute the actual size (type I error), we simulate 300 data sets of 100 time points from Uniform(0,1) distribution in the case when $k = 1$ and 300 data sets of 180 time points from the Uniform(0,1) distribution in the case when $k = 3$. Then the same estimation and testing procedures as those of the power calculation are applied. It appears that both the Wald and LRT tests have very high powers to detect the clusters in the simulated data. The actual sizes of the tests are slightly off if we approximate the distributions of the tests by their large sample asymptotic distributions, indicating the sample sizes used may be still a little too small. But the alternative simulation based Monte-Carlo testing approach is more or less on target. For the particular type of data from model (1), our model based likelihood approaches have more power to detect clusters, and have smaller and much more accurate Type I errors than those of the step regression method proposed by Molinari et al (2001).

In practice, we usually do not know the number of clusters. The AIC or BIC criterion is often used to determine the number of clusters from data that consists of only the time points of events. Table 3 summarizes the model selection results using the AIC and BIC criteria described in Section 4. The numbers reported are consistent in magnitude with those reported in the model selection literature (e.g. Pan, 2001). It appears that the penalty term in the BIC criterion is a little too big and it tends to pick a smaller number of clusters. This phenomena is also reported in additional results provided in Sun (2008).

[Insert Table 3 here]

We carry out an additional study to examine what happens if we use a wrong k to fit the

model. Table 4 lists cluster detection results of using $k = 1, 2, 3$ and 4, while the true number of clusters is $k = 3$. From Table 4, we can see that when using a wrong k ($k = 1$ or 2) less than the true number of clusters $k = 3$, the method almost always pick up one (if wrong $k = 1$) or two (if wrong $k = 2$) of the three true clusters. When using the correct $k = 3$, the method picks up the three true clusters most of the time; this result is consistent with those reported in the second part of Table 2. When using a wrong $k = 4$, a number greater than the true number of clusters $k = 3$, three true clusters appear total 99%, 92.84% and 94.83% times, respectively, and on average 1.1333 clusters are outside the three true clusters. This is evidence that the proposed method picks up three of the true clusters plus one false cluster most of the time.

[Insert Table 4 here]

Finally, we like to remark that we have also carried out simulation studies under an alternative design in which we allow the clusters to change (randomly simulated according to the latent model) in each of the simulation exercises. The simulation results under this design are similar to (only slightly worse than) what we reported here. Due to space limitation, the results are not reported in the paper. See, Sun (2008) for such studies and an interpretation of the results.

7. Discussion

Statistical modeling is one of the most widely used tools in modern applied statistics. A model that mimics the process generating the sample data retrieves information and provides great insight into the problem. We develop such a latent model for a typical cluster generation process. Based on the model, we develop a likelihood inference based detection approach and a Monte-Carlo EM algorithm to identify clusters and estimate cluster locations. Like the generalized scan statistic, the latent modeling approach can flexibly adjust for non-uniform background variation, and can be generalized to two or more dimensions (Sun, 2008). Different from the scan statistic tests, our latent modeling approach has the advantage of not needing to fix the range of cluster sizes or window sizes and number of clusters. Our procedure also gives both a global test of simultaneous clustering and tests of significance of estimated individual clusters. Compared with

the stepwise regression method proposed by Molinari et al (2001) and Demattei and Molinari (2006), our approach shares an important advantage of their approach in its ability to detect multiple clusters of varying sizes in temporal data. But our simulation studies give cases where our procedure has substantially increased efficiency over the stepwise regression approach.

The development in the paper is based on the assumed latent model illustrated in Figure 1. It is possible that there are clusters before the starting time of the study, and the waiting time between the last cluster before time 0, and the first cluster after time 0 is longer than b_1. In the exponential case (with its lack of memory property) the existence of clusters before time 0 does not change the results. In other cases, we may need to model b_1 separately by a truncated $\psi_b(t)$ distribution. It is also possible that either 0 or T fall within a cluster interval. That is, we may assume the first or the last cluster includes 0 or T. Under one of such assumptions, we can in theory use the same techniques described in the paper to develop a similar algorithm. However, for practical purposes, it may be sufficient to use the algorithm outlined in the paper, unless additional information is available. Note that, the probability that an event is exactly at the end point 0 or T is zero; i.e., $P(y_i = 0) = P(y_i = T) = 0$. In practice, without any information outside the given time window $(0, T)$, we can not distinguish a cluster that includes an endpoint with a cluster that starts right after (or ends right before) the endpoint. A simulation study (results not shown here) backs up such an argument.

The likelihood inference described in Section 3.4 is for two-sided tests. In some applications we might be interested in one-sided tests. If the one-sided tests are for each single α_j or in the case that $k = 1$, the Wald type tests and the likelihood ratio test described in section 3.4 can be directly extended to one-sided tests by dividing the p-values in half. For a test that involves multiple α_j's, there is a complication using the likelihood ratio tests. This is inherited from the well known fact that likelihood ratio tests (similar to the F test for multiple regression parameters) are not well suited for one-sided tests of multiple parameters. We also like to note that there are applications in which we would like to limit the parameter space to $\{\alpha_j \geq 1, \text{for } j = 1, 2, \ldots, k\}$

or $\{\alpha_j \leq 1, \text{for}\, j = 1, 2, \ldots, k\}$. In this case, $\alpha_j = 1$ is on the boundary of the parameter space, and the constrained likelihood inference applies; see, e.g., Silvapulle and Sen (2005) and the references therein. Silvapulle and Sen (2005) include an algorithm to compute p-values for constrained likelihood testing problems. We can directly incorporate their algorithm to our problem, with few alterations to the Monte-Carlo EM estimation and cluster location estimation procedures. The theory for constrained likelihood inference is much more complex both in theory and computationally. Since the p-values from constrained likelihood approaches are usually smaller than those from the regular likelihood inference, in practice one may conservatively (with some loss in power) use the two-sided p-values discussed in Section 3.4 for the one-sided testing problems.

Supplementary Materials

The supplementary materials, including the Gibbs algorithms used in Section 3 and the data sets used in Section 6, may be accessed at the Biometrics website http://www.biometrics.tibs.org.

Acknowledgement

The authors thank Dr Molinari for helpful information about their stepwise regression method and providing the R codes for their approach. The authors also thank the reviewers for their constructive suggestions. The research is partially supported by NSF SES 05-18543 and NSA H98230-08-1-0104.

References

Besag J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann Inst Stat Math.* 43:1-59.

Chang, M., Glynn, M.K. and Groseclose, S.L. (2003). Endemic, notifiable bioterrorism-related diseases, United States, 1992 to 1999. *Emerg Infec Dis.* 9, 556-64.

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

Demattei, C. and Molinari, N. (2006). Multiple temporal cluster detection test using exponential inequalities. *Far East Journal of Theoretical Statistics,* 19(2): 231-244.

Demattei, C., Molinari, N and Daures, J.P. (2007). Arbitrarily shaped multiple spatial cluster detection for case event data. *Computational Statistics and Data Analysis.* 51, 3931-3945.

Dembo, A. and Karlin, S. (1992). Poisson approximations for r-scan processes. *Ann. Appl. Probab.* 2, 329-357.

Denison, D. and Holmes, C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics.* 57:143-149.

Diggle, P., Rowlingson, B. and Su, T-L (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics.* 16, 423-434.

Fraiman, R. and Meloche, J. (1999). Multivariate L-estimation (with discussion). *Test.* 8, 255-317.

Fu, J.C. and Lou, W.Y. (2003). *Distribution Theory of Runs and Patterns and its Applications: A Finite Markov Chain Imbedding Approach.* World Scientific.

Gangnon, R.E. and Clayton, M.K. (2000). Bayesian detection and modeling of spatial disease maps. *Biometrics.* 56:922-935.

Gangnon, R.E. and Clayton, M.K. (2003). A hierarchical model for spatially clustered disease rates. *Statistics in Medicine* 22, 3213-3228.

Gangnon, R.E. and Clayton, M.K. (2004). Likelihood-based tests for localized spatial clustering of disease. *Environmetrics.* 15, 797-810.

Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics.* New York: Springer.

Knorr-Held, L. and RaBer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics.* 56:13-21.

Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: Detection and infection.

Statistics in Medicine, 14, 799-810.

Lawson, A.B. (1995). Markov chain Monte Carlo methods for putative pollution source problems in environmental epidemiology. *Statistics in Medicine*. 14, 2473-2486.

Leung, M-Y, Choi, K-P, Xai, A, Chen, L.H.Y. (2005). Non-random clusters of palindromes in Herpesvirus genomes. *J. Computational Biology*. 12, 331-354.

Molinari, N., Bonaldi, C. and Daures, J.P. (2001). Multiple temporal cluster detection. *Biometrics*. 57, 577-583.

Naus, J. and Wallenstein, S. (2004). Multiple window and cluster size scan procedures. *Methodology and Computing in Applied Probability*. 6, 389-400.

Neill, D.B, Moore, A.W., and and Cooper, G.F. (2006). A Bayesian spatial scan statistic. *Advances in Neural Information Processing Systems* 18, 1003-1010.

Pan, W. (2001). Akaike information criterion in generalized estimating equations. *Biometrics*. 57:120-125.

Silvapulle, M.J. and Sen, P.K. (2005). *Constrained Statistical Inference*. Wiley & Sons Inc.

Su, X., Wallenstein, S. and Bishop, D. (2001). Non-overlapping clusters: Approximate distribution and application to molecular biology. *Biometrics*. 57,420-426.

Sun, Q.K. (2008). Statistical models and inferences for detecting multiple temporal and spatial clusters. Ph.D. dissertation, Rutgers University.

Tanner, M. (1993). *Tools for Statistical Inference*. Springer-Verlag: New York.

Wallenstein, S. and Naus, J., (2004). Statistics for temporal surveillance of bio-terrorism. *Morbidity and Mortality Weekly Report*. 53 (Suppl) 74-78.

Waller, L.A., Carlin, B.P., Xia, H., and Gelfand, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates. *J. Am. Statist. Assoc.* 92, 607-617.