Computational Statistics and Data Analysis 53 (2009) 2817–2830

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Surveillance to detect emerging space–time clusters

Renato Assunção [a,*], Thais Correa [b]

[a] Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil
[b] Departamento de Matemática, Universidade Federal de Ouro Preto, Brazil

ARTICLE INFO

Article history:
Available online 5 November 2008

ABSTRACT

The interest is on monitoring incoming space–time events to detect an emergent space–time cluster as early as possible. Assume that point process events are continuously recorded in space and time. In a certain unknown moment, a small localized cluster of increased intensity starts to emerge. Its location is also unknown. The aim is to let an alarm to go off as soon as possible after its emergence, but avoiding that it goes off unnecessarily. The alarm system should also provide an estimate of the cluster location. In addition to that, the alarm system should take into account the purely spatial and the purely temporal heterogeneity, which are not specified by the user. A space–time surveillance system with these characteristics using a martingale approach to derive the surveillance system properties is proposed. The average run length for the situation when there are clusters present in the data is appropriately defined and the method is illustrated in practice. The algorithm is implemented in a freely available stand-alone software and it is also a feature in a freely available GIS system.

1. Introduction

The ongoing and systematic collection of data, as well as its analysis, became essential long ago to the planning, implementation, and evaluation of public health practice. However, as more and more health data are stored in electronic form in a timely way, it is increasing the need for methods which can detect quickly anomalies in a continuously updated database, with minimum input requirements from users. Currently, early disease outbreak detection systems are the object of intense demand by government agencies, specially public health departments. One reason for this heightened interest on the subject is the threat of bioterrorism (Buehler et al., 2004; Henderson, 1999), but these systems have a much larger and much older application scope. In fact, early outbreak detection methods have always been a matter of concern to public health (Hardy, 2001). In the new context of spatially referenced data, these methods face important analytical challenges that include dealing with the adjustment for natural temporal and spatial variation, the unknown time, place and size of an emergent cluster, detecting an outbreak as early as possible, and the lack of suitable population-at-risk data.

Most statistical methods in use for the early detection of disease outbreaks are purely temporal in nature (Sonesson and Bock, 2003; Höhle, 2007; Höhle and Paul, 2008). Hence, they are usually applied to monitor data from large area regions without concern to their geographical location within the monitored regions. They lack power to detect outbreaks that start locally since the affected areas are submerged into large regions with usual incidence rates. One possible solution is to partition the large region into small areas and to apply the purely temporal methods in each small area separately and in parallel. However, this procedure leads to a severe problem of multiple testing, generating many more false signals than the nominal statistical significance level indicate. As a consequence, these purely temporal methods are not appropriate when the data are collected with space and time information. In addition to these problems, there is the expectation that using

* Corresponding author. Tel.: +55 31 3409 5940; fax: +55 31 3409 5924.
E-mail address: assuncao@est.ufmg.br (R. Assunção).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.10.032

the now readily available spatial information can facilitate the detection and localization of emergent clusters (Buckeridge et al., 2005).

The primary purpose of this paper is to suggest a method for the quick detection of space–time emergent clusters in a set of point process events. The requirements to establish a surveillance system, either accounting for spatial structure or not, are generally structured around a basic trade-off: the need for quickly detecting possible outbreaks must be balanced against the need to avoid a high rate of false alarm signals. Our method allows the user to control these trade-off elements in a simple way.

We introduce a stochastic model to describe eventually emerging spatial clusters with minimum requirement of user-defined parameters. When there are no clusters, we assume that the events' density is separable, meaning that it is the product of arbitrary spatial and temporal functions. More importantly, our method does not require the functional specification of these purely spatial or the purely temporal functions. As an alternative to this model, we assume that somewhere, at some moment, one or more space–time high intensity clusters start to emerge. We develop a likelihood model for this pair of hypotheses and monitor the incoming events with a spatial version of the Shyriaev-Roberts statistic. The Shiryaev–Roberts statistic is well known on industrial statistics applications but it is not so common in biometrical work. We use the martingale structure of the Shiryaev–Roberts statistic to derive the values for the tuning parameters of our method.

The next section contains a brief review of the prospective space–time methods available and introduces the main definitions and notation used in the paper. In Section 3, we present our proposal using the Shiryayev-Roberts control chart method based on martingales. Section 4 presents an analysis of the impact of tuning parameters. In Section 5 we show the results of a Monte Carlo study of the method performance. For this, we need to define appropriately what is the expected time until detection when there are clusters present in the data. We are specially interested in the effects of the tuning parameters in our method performance. Section 6 illustrates the use of the method to detect and to identify the particular events that are associated with the space–time clusters. We use the classic Burkitt's lymphoma dataset (Williams et al., 1978) and a Brazilian dataset of Meningitis cases in three years of observation. We analyzed the data using a freely available stand-alone software where our method is implemented. Finally, we close in Section 7 with a discussion and a summary of the main conclusions.

2. Prospective space–time surveillance for localized clusters

The traditional methods for space–time cluster detection are retrospective in nature. That is, they search in a database of past events for evidence of presence of space–time clusters. In contrast, our interest is on prospective methods for geographically restricted: an events' database is updated regularly and then an algorithm should run to help deciding on the emergence of localized space–time clusters. Hence, the clusters must be alive, in the sense that at least some of the most recent events belong to the eventually detected clusters. The regularly updating nature of the database brings two difficult problems. In the first place, the possibility of using too many significance tests as, for example, if one statistical test is carried out every time the database is updated. This induces a severe multiple testing problem with too many false alarms for clusters. As a consequence, such a method would be soon discredited as unreliable. In the second place, reducing in some way the false alarm rate could imply in a long delay to signal a truly emerging space–time cluster. The trade-off between these two problems must be explicitly recognized in any methodology.

A thorough literature review can be found in the book edited by Lawson and Kleinman (2005), in Sonesson and Bock (2003), and in Waller and Gotway (2004). We give here a brief overview of the main proposals. There are non-spatial, purely temporal methods derived from quality control ideas concerned with the monitoring of a stochastic process in time. The Shewart Chart Control is a very simple and popular method but it is not sensitive to small changes in the process. The Cumulative Sum (CUSUM) method accumulates the recent evidence to the previous data until a certain threshold is crossed. It is better than Shewart to detect small changes in the purely temporal process and it has been shown that it has optimal properties in very simple scenarios (see Frisén (2003)). Exponentially weighted moving average also accumulates evidence, as the CUSUM method, but it discounts observations as they get old (Frisén, 2003). All these methods assume that data are independent in time, which is not a realistic assumption in many applications. Kenett and Pollak (1996) uses Shiryaev–Roberts statistics to allow for dependent data. We review this work later in this section.

There are few space–time oriented proposals. One recent promising method has been suggested by Kulldorff (2001) who used a space–time scan statistic for area data. The main difficulty is the control of overall significance level for a sequence of periodic tests, although each individual test has error type I adjusted for all previous analysis at each time moment. We discuss this issue in more detail in the last section.

Rogerson (2001) suggested a statistic based on local Knox statistic that requires only cases data in the form of a space–time point process. Marshall et al. (2007) found severe problems with the probability approximations used in this methods, suggesting that it should not be used. Marshall et al. (2007) demonstrate that the ARL performance of the Rogerson method is highly influenced by some required threshold values, by the population density, and by the region shape. As a consequence, the nominal performance measures associated with Rogerson method are not valid and this makes impossible the tuning of the method without computer simulation.

Kulldorff et al. (2005) have developed a space–time permutation scan statistic for the early detection of disease outbreaks, which is currently in use by the New York City Department of Health for syndromic surveillance. They use a Poisson based

likelihood ratio test statistic scanning over all possible cylinders as clusters candidates. This method does not control overall error type I level for the sequence of periodic analysis. Diggle et al. (2005) and Rodeiro and Lawson (2006) proposed a Bayesian method to model the space–time evolution of the incidence rate and to monitor for changes.

We base our proposal in the Shiryaev–Roberts (SR) surveillance method that was developed only for temporal processes (Shiryaev, 1963; Roberts, 1966; Kenett and Pollak, 1996). Suppose that a sequence of possibly dependent random variables $X_1, X_2, \ldots$ is observed. Two possible models are considered. In one, a sudden change in the stochastic process occurs at the unknown moment k and $f_k(x_1, x_2, \ldots x_t)$ is the joint density distribution of the first t random variables. In the second model, no change ever occurs. In this case, we assume that $k = \infty$ and write $f_\infty(x_1, x_2, \ldots x_t)$ for the joint density. Any surveillance method implies a stopping time T, the first moment when the alarm goes off. Let $E_k(\cdot)$ be the expectation with respect to f_k. The mean $E_\infty(T)$ is called the *Average Run Length* and it is denoted by ARL^0. Clearly, it is desirable to make ARL^0 large. Typically, the user establishes an acceptable minimum threshold B for this parameter. That is, we want $ARL^0 = E_\infty(T) > B$, where B is known.

One approach would be to maximize the likelihood ratio over all possible values of the unknown parameter k defining the statistic

$$\max_{1\le k\le t} \frac{f_k(X_1, \ldots, X_t)}{f_\infty(X_1, \ldots, X_t)}.$$

Rather than adopting this approach, the Shiryaev–Roberts statistic R_t uses the sum of likelihood ratios f_k/f_∞ for all possible change-point moments k:

$$R_t = \sum_{k=1}^{t} \frac{f_{(k)}(X_1, X_2, \ldots, X_t)}{f_\infty(X_1, X_2, \ldots, X_t)}.$$

The alarm goes off if R_t is too large, that is, if $R_t \ge A$. The stopping time T_A is defined as

$$T_A = \min\{t | R_t \ge A\}.$$

It remains to find A such that $ARL^0 = E_\infty(T_A) > B$.

Following the notation of Kenett and Pollak (1996), under P_∞, the sequence

$$\Lambda_{k,t} = \frac{f_k(X_1, X_2, \ldots, X_t)}{f_\infty(X_1, X_2, \ldots, X_t)}$$

is a martingale with expected value equal to 1, even with dependent observations. Therefore,

$$R_t - t = \sum_{k=1}^{t} (\Lambda_{k,t} - 1)$$

is a zero mean martingale. By the Optional Sampling Theorem, we have

$$E_\infty(R_{T_A} - T_A) = 0,$$

and therefore

$$E_\infty(T_A) = E_\infty(R_{T_A}).$$

By definition, $R_{T_A} \ge A$ and hence $E_\infty(T_A) \ge A$. Therefore, taking $A = B$ satisfies the condition $E_\infty(N_B) \ge B$.

There are several advantages associated with the Shiryaev–Roberts method in the time series context. First, it can be shown that it exhibits some optimal properties in some simple scenarios (Pollak, 1985). Pollak (1985) proved that the Shiryaev–Roberts procedure is asymptotically (as $B \to \infty$) optimal in the sense of minimizing the supremum average delay to detection

$$\sup_{k\ge 1} E_k(T - k \mid T \ge k)$$

over all stopping times T that satisfy $E_\infty(T) \ge B$. Yakir (1994) found that this procedure is strictly optimal for the problem of minimizing the average run length to detection over all stopping times T that satisfy $E_\infty(T) \ge B$ when $X_1, X_2, \ldots, X_{k-1}$ are iid random variables. Furthermore, in terms of the delay time for the alarm going off after the purely temporal clusters starts to emerge, the Shiryaev–Roberts and the usual CUSUM method are similar (Shiryaev, 1963; Roberts, 1966; Pollak and Siegmund, 1985; Mevorach and Pollak, 1991). The Shiryaev–Roberts method does not require independence between observations. It can also be shown that Shiryaev–Roberts is at least as efficient as some optimal classical procedures (Kenett and Pollak, 1996).

One difficulty to use the Shiryaev–Roberts method is that it depends on the complete specification of the joint distribution of $X_1, \ldots, X_n$ after a change occurs at k. This is not simple to be done in the purely temporal context and the difficulty increases in the space–time situation. However, we suggest a solution, as we explain next.

3. Detection of emerging space–time clusters

3.1. A model for emerging clusters

Let N be a Poisson process in $\mathbb{R}^3$ partially observed in the three-dimensional region $\mathcal{A} \times (0, \mathcal{T}]$.

The events $(\mathbf{s}_i, t_i) = (x_i, y_i, t_i)$ are indexed by $i = 1, 2, \ldots$, and we assume that $t_1 < t_2 < \cdots$. Let $N(C)$ be the number of events in the set $C \subset \mathcal{A} \times (0, \mathcal{T}]$. We have $N(C)$ distributed as a Poisson random variable with mean $\mu(C)$ given by the integral of the intensity function $\lambda(x, y, t) \geq 0$ over C:

$$\mu(C) = \int_C \lambda(x, y, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t.$$

A special type of set C is a cylinder given by $C = B(\mathbf{s}, \rho) \times (t_a, t_b]$ where $B(\mathbf{s}, \rho)$ is the disc centered at $\mathbf{s} = (x, y) \in \mathcal{A}$ with radius ρ, and $t_a < t_b$.

Let $\mu = \mu(\mathcal{A} \times (0, \mathcal{T}])$ be the expected number of events in the observation region and define the marginal spatial and temporal densities by

$$\lambda_S(x, y) = \mu^{-1} \int_{(0,\mathcal{T}]} \lambda(x, y, t)\mathrm{d}t,$$

and

$$\lambda_{\mathcal{T}}(t) = \mu^{-1} \int_{\mathcal{A}} \lambda(x, y, t)\mathrm{d}x\mathrm{d}y,$$

respectively. Note that

$$\int_{\mathcal{A}} \lambda_S(x, y)\mathrm{d}x\mathrm{d}y = \int_{(0,\mathcal{T}]} \lambda_{\mathcal{T}}(t)\mathrm{d}t = 1.$$

Given that an event $(\mathbf{s}, t)$ occurred in $\mathcal{A} \times (0, \mathcal{T}]$, the functions $\lambda_S(x, y)$ and $\lambda_{\mathcal{T}}(t)$ represent the probability density of $\mathbf{s}$ and t, respectively.

We define now the pair of situations we will consider. The first one is that without space–time clusters. In this case, we have a separable intensity $\lambda(x, y, t) = \mu\, \lambda_S(x, y)\, \lambda_{\mathcal{T}}(t)$ where $\lambda_S(x, y)$ and $\lambda_{\mathcal{T}}(t)$ are arbitrary and unspecified. That is, they are nuisance parameters.

The alternative situation assumes that there exists a time τ, a constant $\varepsilon > 0$, and a cylinder $C = B(\mathbf{s}, \rho) \times (t_a, t_b]$ (yet to be defined) such that

$$\lambda(x, y, t) = \mu\, \lambda_S(x, y)\, \lambda_{\mathcal{T}}(t)\, (1 + \varepsilon\, I_C(x, y, t)).$$

This intensity function can not be written as a product of two functions, one depending only in space and the other only in time. The parameter ε is the relative change on the events intensity within the cluster and it must be specified by the user. Assunção and Maia (2007) and Assunção et al. (2007) used a similar pair of alternative models in the context of space–time point process data.

To define a useful class of cylinders C, we start considering that, if a higher incidence cluster emerges, we must be able to detect it through the observed events. That is, non-events (or void spaces) do not bring information about an emerging cluster. Hence, we decided to constrain t_a to be equal to one of the observed events times t_k. Additionally, the cylinders should be centered around its corresponding spatial location $\mathbf{s}_k$. Since the interest is only on alive clusters, the endpoint t_b is equal to the time t_n of the current last observed event. That is, we consider cylinders of the form $B(\mathbf{s}_k, \rho) \times (t_k, t_n]$, with $k < n$, where $(\mathbf{s}_k, t_k)$ is one previously observed event while $(\mathbf{s}_n, t_n)$ is the last observed event at a given moment. The disc $B(\mathbf{s}_k, \rho)$ has a radius ρ specified by the user. To simplify notation, we denote by $C_{k,n}$ the cylinder $B(\mathbf{s}_k, \rho) \times (t_k, t_n]$ with $k < n$. We extend the notation to include the case $k = n$, writing $C_{n,n}$ to represent the null set.

3.2. A sequential procedure to detect emerging clusters

To define a statistic, we consider the likelihood of the space–time Poisson processes when n events have been observed. If no cluster is emerging, we have

$$L_\infty = \left(\prod_{i=1}^{n} \lambda(x_i, y_i, t_i) \right) \exp\left(- \int_{R^3} \lambda(x, y, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \right).$$

Under the alternative scenario that a cluster started emerging at time $t_k < t_n$, we have

$$L_k = \left(\prod_{i=1}^{n} \lambda(x_i, y_i, t_i)\, (1 + \varepsilon\, I_{C_{k,n}}(x_i, y_i, t_i)) \right) \exp\left(- \int_{R^3} \lambda(x, y, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \right) \exp\left(-\varepsilon \int_{C_{k,n}} \lambda(x, y, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \right),$$

where $\lambda(x, y, t) = \mu\, \lambda_S(x, y)\, \lambda_{\mathcal{T}}(t)$ and $C_{k,n}$ is the putative cluster cylinder.

Therefore, a space–time version of the SR test statistic R_n becomes

$$\begin{aligned} R_n &= \sum_{k=1}^{n} \frac{L_k}{L_\infty} \\ &= \sum_{k=1}^{n} \left\{ \left[\prod_{i=1}^{n} (1+\varepsilon) I_{C_{k,n}}(x_i, y_i, t_i) \right] \exp\left(-\varepsilon \int_{C_{k,n}} \lambda(x, y, t) \mathrm{d}x\mathrm{d}y\mathrm{d}t \right) \right\} \\ &= \sum_{k=1}^{n} (1+\varepsilon)^{N(C_{k,n})} \exp(-\varepsilon\, \mu(C_{k,n})) \qquad (1) \\ &= \sum_{k=1}^{n} \Lambda_{k,n}. \qquad (2) \end{aligned}$$

The expression $\Lambda_{k,n}$ can be seen as a contrast between the observed number $N(C_{k,n})$ of events in $C_{k,n}$ and its expected value under the no-cluster situation. In fact, if ε is small,

$$\Lambda_{k,n} \approx (1+\varepsilon)^{N(C_{k,n})} (1-\varepsilon)^{\mu(C_{k,n})} \approx 1 + \varepsilon \left(N(C_{k,n}) - \mu(C_{k,n}) \right).$$

The parameter $\varepsilon > 0$ is known (user-specified) and measures the anticipated relative change in the events' density. Note that

$$\Lambda_{n,n} = (1+\varepsilon)^{N(C_{n,n,})} \exp(-\varepsilon \mu(C_{n,n})) = (1+\varepsilon)^0 \exp(0) = 1.$$

Our surveillance method calculates R_n as the n-th event arrives, substituting the unknown $\mu(C_{k,n})$ by an estimate $\hat{\mu}(C_{k,n})$. The estimation of $\mu(C_{k,n})$ is discussed in Section 3.3. The alarm goes off when $R_n \geq A$ for the first time.

In summary, the algorithm associated with our proposal requires as input:

- a set of n case events, specified by their spatial coordinates x, y and time t;
- the value of three user-specified tuning parameters:
  . the anticipated relative change ε in the density within the cluster;
  . the anticipated radius ρ for the cluster;
  . the threshold A, which should be approximately equal to the desired ARL^0.

Iteratively in n, we calculate R_n. The output is a sequence of values R_n where n is the number of events. If $R_n > A$ for any n, the alarm goes off.

If the alarm goes off, one important practical issue is the space–time cluster location. Suppose that the alarm goes off at the n-th event. That is, $R_t \geq A$ for the first time at $t = n$. Since

$$R_n = \sum_{k=1}^{n} \Lambda_{k,n} = \Lambda_{1,n} + \Lambda_{2,n} + \ldots \Lambda_{n,n}.$$

The large values of $\Lambda_{k,n}$ are those contributing to the alarm triggering. Let $\Lambda_{k*,n} = \max\{\Lambda_{k,n},\, 1 \leq k \leq n\}$. A cluster estimate is built by taking the spatial coordinates (x_k*, y_k*) of the $k*$-th event as the center of the cylinder basis, and by taking its height equal to $[t_k*, t_n]$. That is, $C_{k^*,n}$ is the space–time cluster estimate.

3.3. Estimation of $\mu(C_{k,n})$

Typically, the user will not be able to specify the purely spatial $\lambda_S(x, y)$ and the purely temporal $\lambda_T(t)$ functions. Therefore, it is relevant in practice to alleviate him from such requirements. Rather than using the mean $\mu(C_{k,n})$ in (1), we use the data themselves to estimate it. From the non-homogeneous Poisson process properties, under the null hypothesis, for $k < n$ we have:

$$\mu(C_{k,n}) = \int_{C_{k,n}} \lambda(x, y, t) \mathrm{d}x\mathrm{d}y\mathrm{d}t = \mu \int_{B(\mathbf{s}_k, \rho)} \lambda_S(x, y) \mathrm{d}x\mathrm{d}y \int_{(t_k, t_n]} \lambda_T(t) \mathrm{d}t.$$

Therefore, an estimate of $\mu(C_{k,n})$ under the hypothesis that there are no clusters is given by

$$\hat{\mu}(C_{k,n}) = \frac{N(B(\mathbf{s}_k, \rho) \times (0, t_n])\, N(\mathcal{A} \times (t_k, t_n])}{n},$$

where $N(B(\mathbf{s}_k, \rho) \times (0, t_n])$ is the number of events within the disc $B(\mathbf{s}_k, \rho)$ irrespective of occurrence time, $N(\mathcal{A} \times (t_k, t_n])$ is the number of events between times t_k and t_n, irrespective of their spatial location, and n is the total number of events at that moment (see Fig. 1).

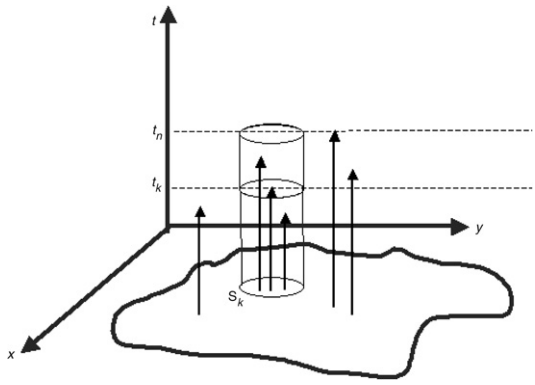

Fig. 1. The estimate $\hat{\mu}(C_{k,n})$.

3.4. Iterative calculation of R_n

Every new event requires the recalculation of all n terms in (2). We can decrease substantially the amount of numerical calculations by means of an iterative procedure. For $k < n$, let $I_{k,n} = I(\|\mathbf{s}_n - \mathbf{s}_k\| \leq \rho)$. We have

$$N(C_{k,n+1}) = N(C_{k,n}) + I_{k,n+1}$$

and

$$\mu(C_{k,n+1}) = \mu(C_{k,n}) + \mu\left(B(\mathbf{s}_k, \rho) \times (t_n, t_{n+1}]\right).$$

Therefore, since $\Lambda_{1,1} = 1$, we can write recursively for $k < n + 1$,

$$\Lambda_{k,n+1} = \Lambda_{k,n}\,(1+\epsilon)^{I_{k,n+1}}\,\exp\left(-\epsilon\mu\left(B(\mathbf{s}_k, \rho) \times (t_n, t_{n+1}]\right)\right).$$

By definition, $\Lambda_{n+1,n+1} = 1$ and this completes the recursion.

In practice, to run the procedure, we need $\hat{\mu}(C_{k,n+1})$, rather than $\mu(C_{k,n+1})$. However, this estimated term can also be calculated iteratively for $k < n + 1$:

$$\begin{aligned}
\hat{\mu}(C_{k,n+1}) &= \frac{1}{n+1} N(B(\mathbf{s}_k, \rho) \times (0, t_{n+1}]) N(\mathcal{A} \times (t_k, t_{n+1}]) \\
&= \frac{n+1-k}{n+1}\left(N(B(\mathbf{s}_k, \rho) \times (0, t_n]) + I_{k,n+1}\right) \\
&= \frac{n}{n+1}\,\frac{n-k+1}{n-k}\,\hat{\mu}(C_{k,n}) + \frac{n+1-k}{n+1}\,I_{k,n+1} \\
&= \hat{\mu}(C_{k,n}) + \frac{k}{(n+1)(n-k)}\hat{\mu}(C_{k,n}) + \frac{n+1-k}{n+1}\,I_{k,n+1}.
\end{aligned}$$

For $k < n + 1$, we can write

$$\hat{\Lambda}_{k,n+1} = (1+\varepsilon)^{N(C_{k,n})+I_{k,n+1}} \exp\left(-\varepsilon J_{n,k}\right) + \hat{\Lambda}_{n+1,n+1},$$

where

$$J_{n,k} = \frac{n(n-k+1)}{(n+1)(n-k)}\hat{\mu}(C_{k,n}) + \frac{n+1-k}{n+1} I_{k,n+1}.$$

Therefore,

$$\begin{aligned}
R_{n+1} &= \sum_{k=1}^{n+1} \hat{\Lambda}_{k,n+1} \\
&= 1 + \sum_{k=1}^{n} \hat{\Lambda}_{k,n} \exp\left(\frac{-\varepsilon\, k}{(n+1)(n-k)}\,\hat{\mu}(C_{k,n})\right) L_{\varepsilon,n,k},
\end{aligned}$$

where

$$L_{\varepsilon,n,k} = \left((1+\varepsilon)\exp\left(-\varepsilon\,\frac{n+1-k}{n+1}\right)\right)^{I_{k,n+1}}.$$

We let $1 = \Lambda_{n+1,n+1} = \hat{\Lambda}_{n+1,n+1}$. As a consequence, we obtain R_{n+1} simply by updating the values of $\hat{\Lambda}_{k,n}$ with a few numerical calculations.

4. Choice of tuning parameters

The variance of the test statistic $R_n = \sum_k \Lambda_{k,n}$ increases with ε when we have no clusters. More importantly, the distribution of R_n is quite asymmetric. Indeed, for one side a large negative deviate of $N(C_{k,n})$ from its mean $\mu(C_{k,n})$ push $\Lambda_{k,n}$ towards its lower bound, equal to zero. For the other side, increasing a large positive deviate drives $\Lambda_{k,n}$ towards infinity. As a consequence, the false alarm rate increases with ε. The simulations in Section 5 will show these effects of changing ε on the surveillance system performance. Although analytical results are difficult to obtain, simple approximations provide some insight into this trade-off.

When there is no cluster emerging, we have $E(\Lambda_{k,n}) = 1$ for all ε because

$$\begin{aligned} E(\Lambda_{k,n}|H_0) &= \exp(-\varepsilon\,\mu(C_{k,n}))\,E[(1+\varepsilon)^{N(C_{k,n})}] \\ &= \exp(-\varepsilon\,\mu(C_{k,n}))\,\exp(\mu(C_{k,n})(1+\varepsilon))\exp(-\mu(C_{k,n})) \\ &= 1. \end{aligned}$$

Then, $E(R_n) = n$ for all ε, as we could expect from the martingale approach of Section 2.

Suppose now that a cluster emerges and that $N(C_{k,n}) \sim \text{Poisson}(\mu(C_{k,n})\,(1+\varepsilon^*))$. At this point, we distinguish the true relative change ε^* from the one specified by the method (ε). They do not need to coincide. In this alternative situation we have

$$\begin{aligned} E(\Lambda_{k,n}) &= \exp(-\varepsilon\,\mu(C_{k,n}))\,E[(1+\varepsilon)^{N(C_{k,n})}] \\ &= \exp(-\varepsilon\,\mu(C_{k,n}))\,\exp[\mu(C_{k,n})(1+\varepsilon^*)(1+\varepsilon-1)] \\ &= \exp[\mu(C_{k,n})\varepsilon\varepsilon^*] > 1. \end{aligned}$$

That is, $E(\Lambda_{k,n})$ increases with ε (and with ε^*).

Therefore, apparently, the choice of ε does not affect R_n when there is no cluster. When a cluster emerges, choosing a large ε^* will increase R_n and speed up the threshold crossing, as we wish in this case. Hence, it seems that taking $\varepsilon \to \infty$ is the a good strategy.

However, there is a penalty for choosing ε^* too large in that the $\text{Var}(\Lambda_{k,n})$ increases with ε^* when there is no cluster. In fact,

$$\text{Var}(\Lambda_{k,n}) = \text{Var}[\exp(-\varepsilon^*\mu(C_{k,n}))\,(1+\varepsilon^*)^{N(C_{k,n})}]$$

which is equal to

$$\exp(-2\varepsilon^*\,\mu(C_{k,n}))\,\{\exp(\mu(C_{k,n})((1+\varepsilon^*)^2-1)) - \exp(2\mu(C_{k,n})(1+\varepsilon^*-1))\}.$$

This reduces to

$$\exp(\mu(C_{k,n})\varepsilon^{*2}) - 1 \to \infty,$$

as $\varepsilon^* \to \infty$.

Ultimately, this will increase the false alarm rate. Under no cluster, increasing ε^* too much implies in a larger variability of R_n. This is so, because the pairs of terms $\Lambda_{k,n}$ are either uncorrelated (if the corresponding cylinders are non-intersecting) or positively correlated (with correlation proportional of the intersecting volume). As a consequence, the variance of the stopping time T_A increases as well as the probability that R_{T_A} crosses the threshold A at a very early moment, as well as much later than the expected $E(T_A) = A$. Hence, there is a larger probability that the threshold will be crossed before any cluster emergence. These effects will be clear in the simulations of Section 5.

As shown in the simulations, changing the radius produces small effects on the surveillance system performance.

5. Method performance

We evaluated the performance of our method with Monte Carlo simulation using three types of scenarios. In the first type, there were no clusters and the purpose is to evaluate if the approximation $A = B \approx ARL^0$ is appropriate. The geographical region was the rectangle $[0, 10] \times [0, 10]$ and the spatial location of an event was obtained by independently and uniformly generating coordinates on the square. Times between events were modelled by independent exponential random variables with mean equal to 1. Times and locations were independently generated and hence $\lambda(x, y, t) = 0.01$. The events were generated sequentially until the alarm was triggered.

In the second type of scenario, in addition to the events generated as in the first scenario, we also simulated events within a cylindrical cluster that emerged at some moment. The purpose of this second type of scenario is to evaluate the detection performance and the effects of the user-specified tuning parameters of our method. The cluster had the square basis $[4, 5] \times [4, 5]$. It was kept alive from its outbreak until the alarm went off. We selected three different times for the cluster outbreak. In one case, the cluster starts at the beginning of the time period (when $t = 50$) and we label this as *case B*. In the second case, the cluster starts in the middle of the time period, when $t = 150$, and this is labelled *case M*. Finally, in the third case, the cylinder cluster starts late, at $t = 300$, and we labelled this as *case L*.

In each cylinder cluster, we have two types of events: those generated in the larger square and that happened to fall within the cluster, and those events generated within the cluster itself. The intensity at a location (x, y, t) within the cluster is slightly larger than 1.2 times the intensity outside the cluster. This implies that the correct value of the tuning parameter is approximately $\varepsilon = 0.2$.

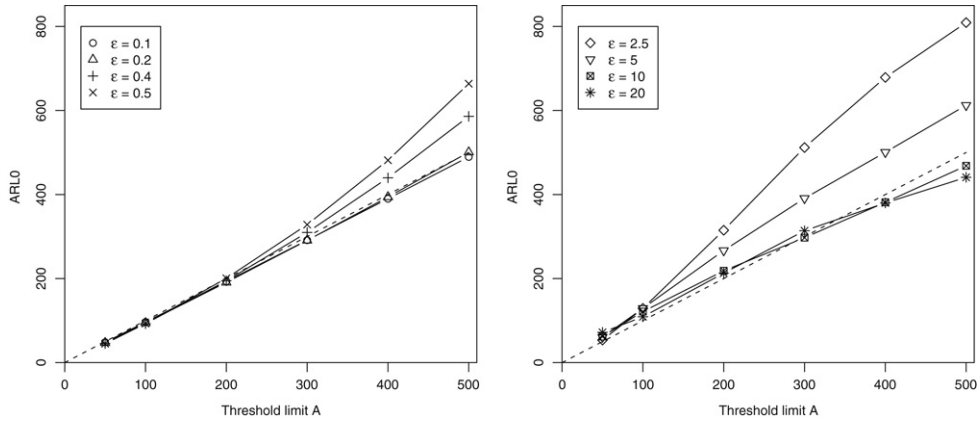

Fig. 2. Scenario without cluster. The plots show the estimated $ARL^0 = E(T_A)$, the average number of events observed before a false signal is issued, versus the threshold limit A. Each curve corresponds to a value of ε. In all cases, we used $\rho = 1$.

We will call this second type of scenario the homogeneous scenario because there is no spatial variation in the events' intensity except that due to the cluster emergence. The third type of scenario was generated in the same way as the second type except by the use of a spatially heterogeneous density of the events locations. In this third scenario type, the spatial coordinates were generated using a mixture of four bivariate normal distributions. Therefore, at any time, some regions were more likely to observe an event than other regions.

We generated 1000 independent replications of each scenario. To run the surveillance procedure, we used several values for ε: 0.1, 0.2, 0.4, 0.5, 2.5, 5, 10, 20. Note that, for the scenario without clusters, the true value of this parameter is zero, while in the scenarios with clusters, it is equal to 0.2.

No comparison with the scan statistic method from Kulldorff (2001) has been attempted because it takes too long to run. In each scenario, for a single simulated space time dataset with 1000 events, Kulldorff's method takes 2 h and 26 min in a machine with 1.66 GHz and 1 Gb RAM when we ask for 999 Monte Carlo replications. To repeat this procedure for all the simulated datasets over the many scenarios is unfeasible.

5.1. Scenario without clusters

We used the values 50, 100, 200, 300, 400, and 500 for the threshold limit A and $\rho = 1$ for the spatial radius of the presumed cluster. If there are no clusters, any surveillance signal is a false signal. If the time period goes to infinity, the statistic R_n will cross the threshold A with probability one, irrespective of the presence of a real cluster. It is expected that, in a fixed time period, the number of events until the alarm goes off increases with the increase of the threshold A. If the approximation $A = B$ is reasonable, we should have $A \approx ARL^0$.

The graph in Fig. 2 shows in the vertical axis the estimated ARL^0, the average number of events observed before a false signal is issued, versus the threshold limit A in the horizontal axis. The dashed line represents the line $ARL^0 = A$. The other lines represent the test results for different values of ε.

As we expect, ARL^0 increases with the threshold A. For $\varepsilon = 0.1$ and 0.2, the approximation $ARL^0 = A$ is very good. For other values of ε this approximation deteriorates with the increase of A. Initially, the departure from the line $ARL^0 = A$ is larger the greater is ε. For example, when $A = 500$ and $\varepsilon = 0.5$, the estimated ARL^0 is 33% larger than the nominal value of 500. However, the difference between the nominal value A and the true ARL^0 does not increase monotonically with ε. With $\varepsilon \geq 10$, this difference is almost null. Setting $\varepsilon \geq 10$ is an extreme choice since it means an emerging cluster with intensity 10 times larger than the baseline intensity. It is unlikely that our method is envisioned for such type of anticipated changes.

This result shows that, when no cluster is present and within practical bounds for the expected relative change in intensity, selecting larger values for the user-specified ε parameter leads to conservative ARL^0. That is, if the user is unsure about ε, selecting a larger value leads to longer ARL^0 times than the nominal value A.

However, there is a trade-off in selecting ε too large in that the false alarm rate can increase, as we show in the next section. The reason for this behavior is the increase on the standard deviation of the stopping time T_A with the increase of ε. Table 1 shows this standard deviation for different choices of ε and threshold A. The larger variability will increase the chances of observing much earlier and much later T_A than its expected value. This increases the chance that an unmotivated alarm sounds off before any cluster emerges. This behavior can be fully understood after we present the results of simulations with clusters.

5.2. Scenarios with clusters

One of the usual performance measures of temporal surveillance systems is the *CED*, the expected delay to trigger a signal after the emergence of a cluster. Assume that the cluster emerges at τ. Given that the stopping time T_A is greater or equal

Table 1
Standard deviation of the stopping time T_A for different choices of threshold A and ε. In all cases we used $\rho = 1$.

Threshold	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.4$	$\varepsilon = 0.5$	$\varepsilon = 2.5$	$\varepsilon = 5$	$\varepsilon = 10$	$\varepsilon = 20$
50	0.498	0.518	0.928	1.143	7.115	12.610	17.473	23.009
100	0.561	0.996	1.933	2.483	26.050	39.971	40.410	40.037
200	1.051	2.088	4.746	6.630	95.582	103.240	96.583	97.584
300	1.565	3.291	8.851	13.663	180.976	178.143	142.034	163.956
400	2.137	4.647	15.484	26.544	254.284	233.516	196.761	210.668
500	2.982	6.915	25.488	48.344	294.398	283.244	239.812	254.411

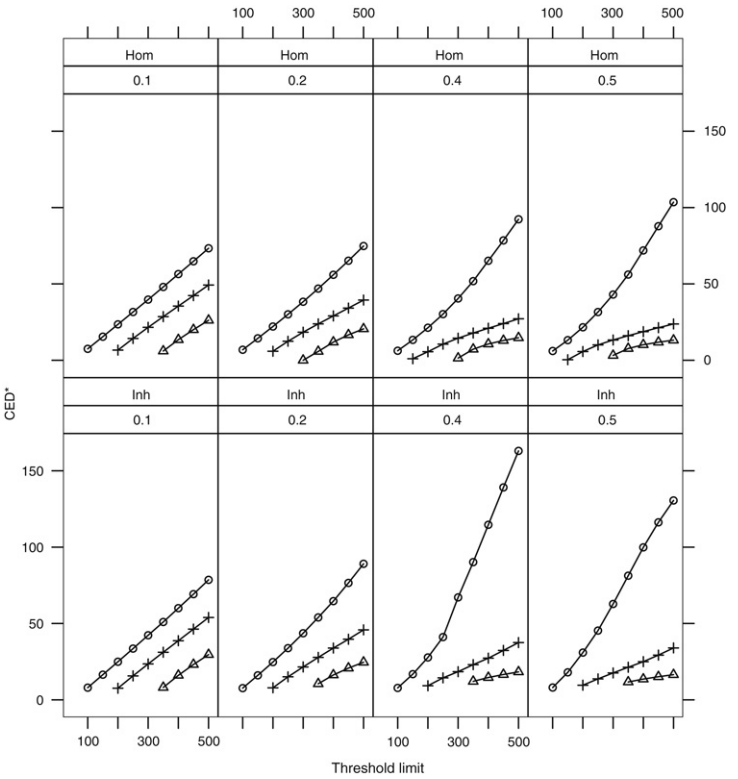

Fig. 3. Estimated $CED^*(\tau)$ against the values of the threshold A. The first row of plots corresponds to the homogeneous scenarios (Hom) while the second row corresponds to the inhomogeneous scenarios (Inh). The different columns correspond to different values of ε. Each plot has three lines. The circles correspond to case B, when the cluster emerges soon in the observation period ($\tau = 50$). The crosses correspond to case M ($\tau = 150$), and the triangles correspond to case L ($\tau = 300$).

to τ, the usual definition of CED is the expected number of observations one needs to wait until the alarm signal. That is, $CED(\tau) = E[T_A - \tau | T_A \geq \tau]$. One problem with this definition in the space–time situation is that all events between τ and T_A contribute to CED, either they belong to the cluster or not. We think that a more appropriate CED definition is the average number of events within the space–time cluster until the alarm goes off. We denote this measure by CED^* to distinguish it from the more usual temporal definition.

We tested with four different values for the spatial radius ρ: 0.25, 0.5, 1, and 2. It is worth remembering that the cluster basis is an unit square. We considered a more refined grid for values of the threshold A than in the without cluster scenario. Namely, we set A equal to 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500.

The plots in Fig. 3 show the estimated $CED^*(\tau)$ against the values of the threshold A. The first row of plots corresponds to the homogeneous scenarios (Hom) while the second row corresponds to the inhomogeneous scenarios (Inh). The different columns correspond to different values of ε. There is one plot for each value of the anticipated relative intensity increase ε. In each plot, we have three lines depending on the value of τ, which can be $\tau = 50$ (case B), $\tau = 150$ (case M), and $\tau = 300$ (case L). It only makes sense to analyze the data for threshold values larger than the emerging cluster time τ. Accordingly, for case B, we do not show $CED^*(\tau)$ when $A = 50$ because almost always the alarm is falsely motivated. The same occurs in case M with threshold $A \leq 150$, and in case L with threshold $A \leq 300$. The estimated $CED^*(\tau)$ in these plots are obtained in simulations where the false alarm percentage is zero.

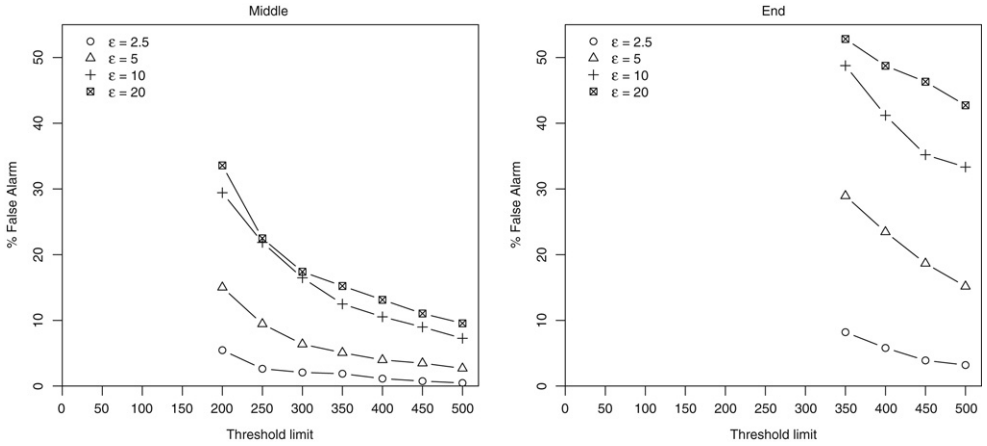

Fig. 4. Left hand side: False alarm rates versus threshold limit A for the scenarios when the cluster emerges on the middle of the observation period with $\varepsilon = 2.5, 5, 10, 20$. Right hand side: Idem for cluster emerging at the end of the observation period.

We analyze initially only the the Middle and End scenarios. In these cases, the plots in Fig. 3 lead to the preliminary conclusion that the larger the ε, the better the performance of the surveillance method. Indeed, the $CED^*(\tau)$ decreases with the increase of ε. One should rather set $\varepsilon = 0.5$ than $\varepsilon = 0.2$, which is the true value used in the simulation.

However, increasing ε too much leads to very large false alarm rates. Fig. 4 shows the false alarm rates for ε equal to 2.5, 5.0, 10, and 20 in the homogeneous scenarios. The left hand size plot corresponds to the Middle and the right hand side to the End scenario. Then, it is clear that increasing ε without bounds renders the method useless because the false alarm rate is beyond tolerable standards. This behavior is virtually identical in the inhomogeneous scenarios.

Returning to the more practically oriented values of ε shown in Fig. 3, we see that the delay for the alarm to go off is smaller if the cluster is at the end of the observation period. Basically, this means that the alarm system learns what is the intensity for a long time under the no cluster situation. When a cluster finally starts emerging, it is quickly detected. Unless the cluster is at the beginning of the observation period, the $CED^*(\tau)$ curves are almost parallel lines. Hence, the effect of the emerging time τ in the $CED^*(\tau)$ is approximately linear, unless τ is close to zero.

There is a heavy penalty for clusters located at the beginning of the observation period and with ε larger than the true relative change in intensity. In this situation, the expected delay increases very quickly. The alarm system takes a very long time to tell apart what is the baseline intensity from the emerging cluster higher intensity.

Contrasting the homogeneous with the inhomogeneous case, we can see from Fig. 3 that, in the inhomogeneous case, the $CED^*(\tau)$ is greater than in the homogeneous case. This effect is especially dramatic if the cluster is located in the beginning of the observation period and ε is larger than the true relative change.

In Fig. 5, we show the effects of the radius ρ on the surveillance system performance for the homogeneous case. We ran simulations of the homogeneous scenario with spatial radius $\rho = 0.25, 0.5, 1$, and 2. The inhomogeneous case has virtually identical conclusions and it is not shown. The relative change parameter ε had the values 0.1, 0.2, 0.4, and 0.5. With respect to the observation period, the clusters started early ($\tau = 50$), at the middle ($\tau = 150$), or late ($\tau = 300$). Most often, the procedure is insensitive to the choice of the radius ρ. Except for the cluster at the beginning and large ε, there is very little difference on the estimated $CED^*(\tau)$. The exceptional behavior occurs only in a special conjunction of factors: when the parameter ρ is larger than the true cluster radius, when the clusters starts emerging early, and when ε is much larger than its true value.

6. Illustrative examples

We illustrate our method using two real datasets: the Burkitt's lymphoma cases in Uganda, the same dataset used in Rogerson (2001), and the *Meningitis* cases occurred in Belo Horizonte, Brazil, between 2001 and 2005.

6.1. Burkitt's lymphoma cases in Uganda

A classical example of retrospective detection of space–time clustering is that based on the Burkitt's lymphoma in Uganda (Williams et al., 1978). The data consist of the place of residence and onset time for all 188 cases of Burkitt's lymphoma between 1961 and 1975 in the West Nile district in Uganda (see Fig. 6). Rogerson (2001) found evidence of space–time clusters using local Knox tests and adopting a probability of false alarm of 0.1. Notwithstanding the problems found in Rogerson's method by Marshall et al. (2007), we compare our results with his.

The tuning parameters in our surveillance method were: $\varepsilon = 0.1, 0.2, 0.4$, and 0.5; $\rho = 2.5, 5, 10$, and 20 km; alarm threshold $A = 161$. Hence, in average, we expect 161 events before the alarm goes off falsely. In all cases, the alarm went off and the triggering event varied from 142 to 158. The difference is $158 - 142 = 16$, which corresponds to 8.5% of the total

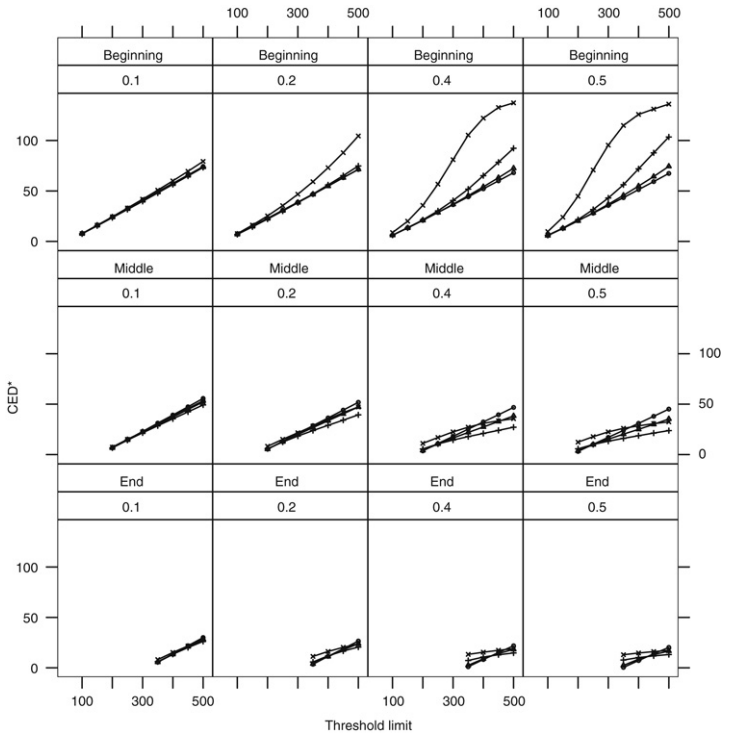

Fig. 5. Effect of changing ρ. The rows correspond to the three cluster emerging time $\tau = 50, 150, 300$, and the columns correspond to different values of ε. Only the homogeneous case was considered. The curves with circles correspond to $\rho = 0.25$, the curves with triangle to $\rho = 0.5$, the curves with crosses to $\rho = 1.0$, and the curves with axes to $\rho = 2.0$. The true value of ρ is 0.5.

number of observations. With respect to the emerging time τ, the estimates varied from 103 to 138. The largest number corresponds to the smallest radius (2.5). Except for this rather extreme radius, all the estimated starting times τ varied from 103 to 107, a very short range. More relevant than this, the estimated clusters are located approximately in the same region in all cases. Hence, the results are relatively insensitive to the tuning parameters values.

Fig. 6 shows the graphs of R_n versus n for ρ equal to 10 and 20 km. For $\rho = 20$ km and $\varepsilon = 0.5$, the alarm goes off at event number 148 (February, 1973) and the method estimates that the clusters started on event 107 (November, 1970). This cluster is represented in the map on the right hand of Fig. 6. Note that, among the 40 events occurring all over the map between November, 1970 and February, 1973, only 20 belong to the cluster.

It is notable that the cases in the emerging cluster shown in Fig. 6 coincide with one of the clusters identified by Williams et al. (1978) through the pairs of cases whose disease onset was in the period 1972–1973, within 10 km apart and 180 days of each other.

Rogerson (2001) ran his procedure with several different tuning parameters and his results are sensitive to these choices. Many of his detected clusters coincide with ours. However, one should keep in mind the criticisms to Rogerson procedure presented in Marshall et al. (2007).

We ran the space–time prospective scan statistic method proposed by Kulldorff (2001) and implemented in the software SaTScan (Kulldorff, 2006). SaTScan is a freely available software and it was developed under the joint auspices of Martin Kulldorff, the National Cancer Institute, and Farzad Mostashari of the New York City Department of Health and Mental Hygiene.

There are difficulties comparing our method with that of Kulldorff (2001). His method, as implemented in the software, receives a dataset and scans for an alive cluster. That is, it searches for a cylinder-shaped cluster whose height ends at the last available observation time. Running his method with all 188 events will not allow for clusters that could have been found before the last observation. To avoid running his method manually repeatedly by adding a single observation each time, we ran it initially with all 188 observations. The scan statistic finds a non-significant cluster centered at the spatial coordinates (273, 332), with a radius equal to 13.42 km, and ranging from 9/2/1972 to 10/24/1975, when the scan alarm sounded off. Based on 999 simulations, the Monte Carlo p-value associated with this space–time cluster is equal to 0.129.

We then used this same 13.42 km radius found by the scan statistic in our own method. The alarm sounded off four times, in 6/15/1973, 4/24/1973, and 2/1/1973, for ε equal to 0.1, 0.2, 0.4, and 0.5, respectively. The clusters found were all centered at the spatial coordinates (273, 332), the same non-significant cluster found by the scan statistic based on all data points. Our method estimated the cluster emergence at 31/3/1971.

Finally, we ran Kulldorff's method again but using only the events that happened until the dates our method sounded off. Using the events until the endpoint 6/15/1973 or until the endpoint 4/24/1973, we have essentially the same results

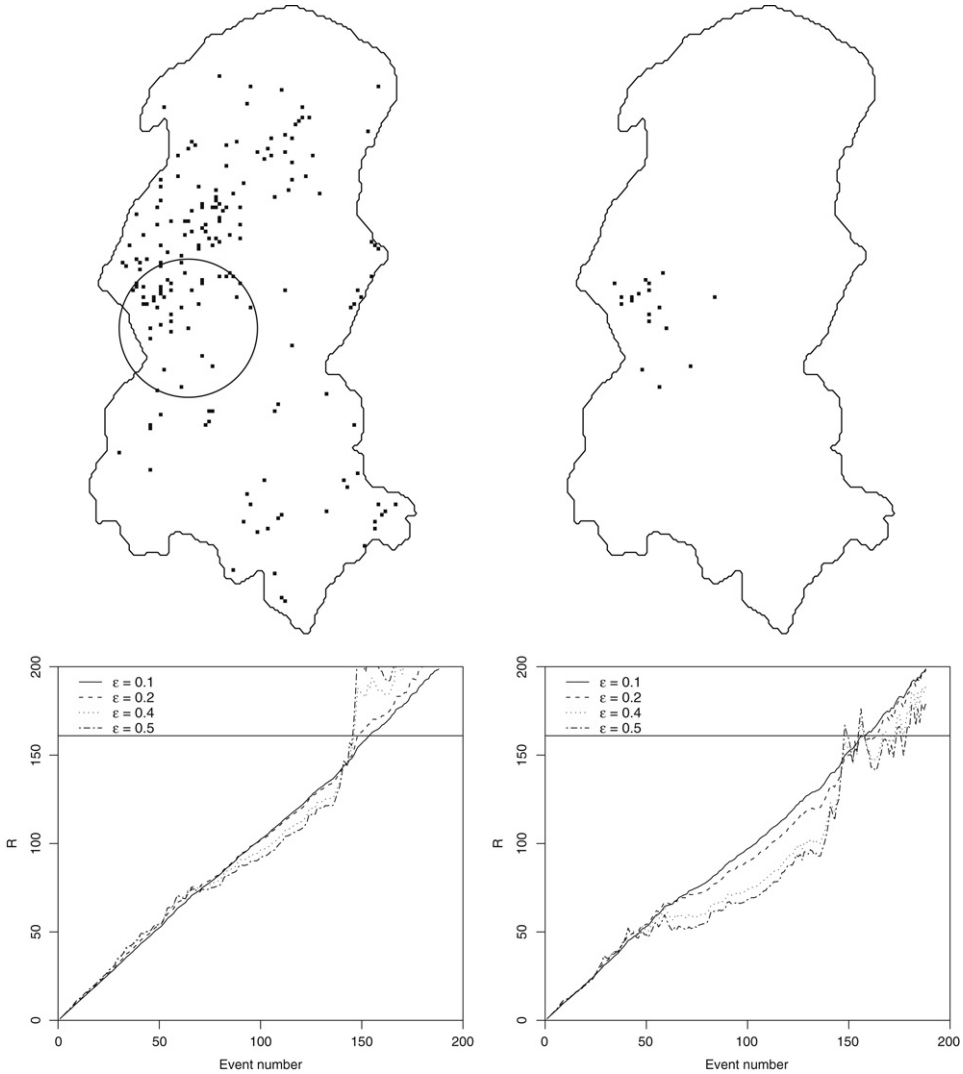

Fig. 6. Burkitt's lymphoma cases in West Nile district of Uganda from 1961 to 1975 (study region is approximately 80 km × 170 km). The left hand side map shows all the events in the period while the right hand side map shows the events identified in the emerging cluster by our method. Each one of the plots shows R_n versus n for four different choices of ε: 0.1, 0.2, 0.4, and 0.5. The left hand side plot uses $\rho = 10$ km and the right hand side plot uses $\rho = 20$ km.

as our own method. A 5% significant cluster is found with radius equal to 11.40 km centered at (273, 332), and ranging from 2/9/1972 until the corresponding endpoint. Running Kulldorff's procedure with the events until 2/1/1973, we find a 5% significant cluster with radius equal to 5.0 km centered at the spatial coordinates (263, 333), and emerging at 9/2/1972, the same date as the previous cases. Therefore, in this example, the two methods give similar results.

6.2. Meningitis cases in Belo Horizonte

We apply our method using data with place and onset date for 1001 Meningitis cases that occurred between 2001 and 2005 in Belo Horizonte, Brazil (see Fig. 7). In 60% of the days no cases were recorded and, in 72% of the remaining days, only one case was recorded. The maximum number of cases in a single day was equal to 6. From the time series plots in Fig. 7, no discernible trend or periodicity is present.

The tested the following parameters: $\varepsilon = 0.1, 0.2, 0.4$, and 0.5; $\rho = 1, 2, 3$, and 4 km; alarm threshold $A = 500$. Threshold $A = 500$ means that we expect 500 cases before the alarm goes off without need. Since we have around 200 cases per year, we are expecting two unmotivated alarms in a period of 5 years.

The alarm went off in all situations, irrespective of the parameters values. The triggering event varied from 490 to 949 and it increased if either of the tuning parameters, ε and ρ, increased. However, there was a positive interaction between these two tuning parameters: the waiting time increased faster with ρ if ε was larger.

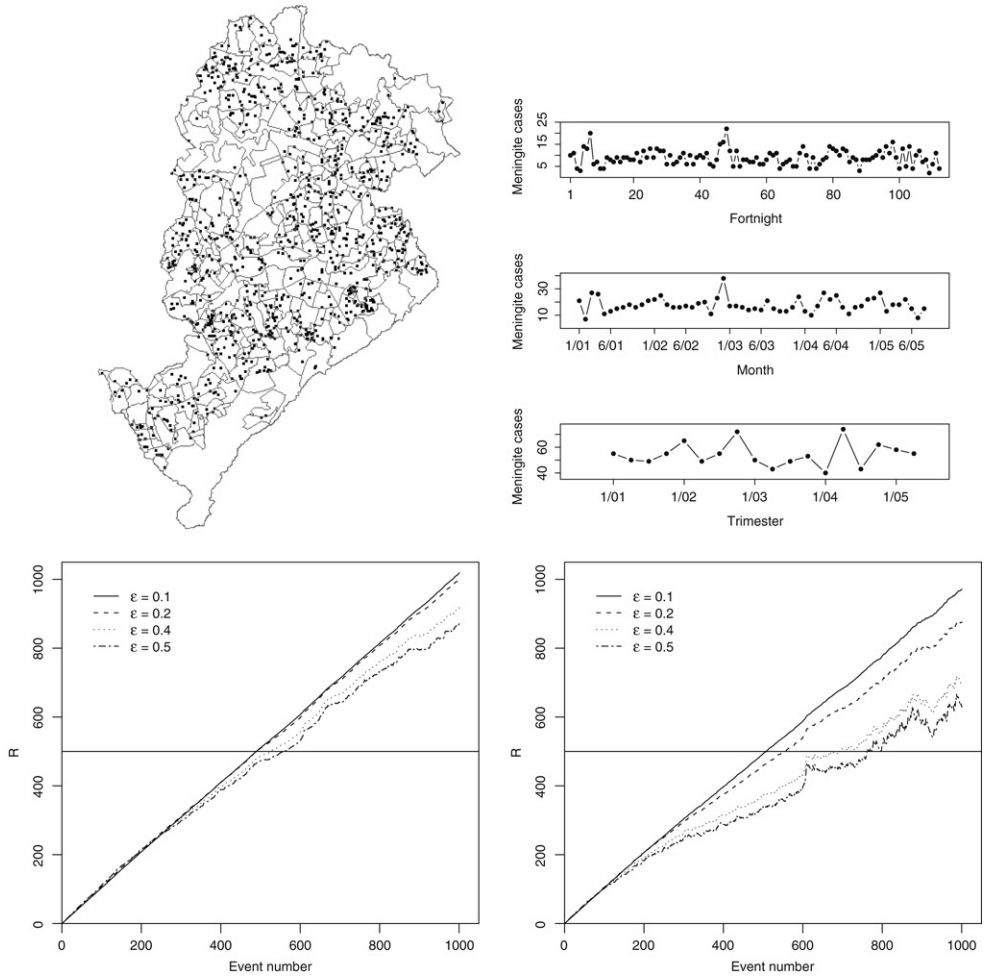

Fig. 7. Map of Belo Horizonte divided into neighborhoods and the location of 1001 Meningitis cases that occurred between 2001 and 2005. The study region has approximately 31 km × 16 km. The time series plots show the number of events in different time units: fortnightly, monthly, and quarterly. Each one of the plots shows R_n versus n for four different choices of ε: 0.1, 0.2, 0.4, and 0.5. The threshold is $A = 500$. The left hand side plot uses $\rho = 1$ km and the right hand side plot uses $\rho = 2$ km.

Fig. 7 shows the R_n statistics (2) versus n for $\varepsilon = 0.1, 0.2, 0.4, 0.5$ and $\rho = 1.0, 2.0$ km. These plots illustrate the effects of changing ρ and ε.

City health officials were not suspicious of any emerging cluster during this period before our analysis. Since the triggering event occur around the expected time under the no cluster situation, our method supports this opinion. There is no convincing evidence for the emergence of meningitis clusters in Belo Horizonte.

This lack of evidence is reinforced by Kulldorff's prospective scan statistic method. Running his method with all the events, we did not find a significant cluster. Its most likely cluster, with p-value equal to 0.129, had radius zero, including only two events with identical spatial coordinates.

7. Conclusions

Compared to the main space–time surveillance methods in the literature, our method has some advantages and disadvantages. Kulldorff (2001) does not requires tuning parameters and he does not use the concepts of average run length and conditional expected delay preferring the hypothesis testing concepts of error type I and power. One problem with his method is that it does not control the error type I over repeated and periodic surveillance. It adjusts for all previous analysis in each time moment the scan is performed and a correct α error type I probability is achieved at that moment. However, considering the simultaneous inference for all points in time, the correct level is not α. This is clearly seen if one considers the situation in which no cluster ever emerges. Running a α level test at each moment in sequence, even when controlling for the past analysis at each moment, we are bound to provide a significant result eventually. Therefore, the true significance level would be 1 for the infinite sequence of tests.

We avoid this problem by adopting the quality control ideas of average run length but our method requires the specification of tuning parameters. Although we think that users should be able to propose reasonable values for these parameters, one needs more studies to understand fully the impact of them in practice. This need to set tuning parameters is also a requirement in Rogerson (2001) but, as Marshall et al. (2007) shows, his method has several shortcomings.

Our surveillance method assumes a spatially circular shaped cluster. A completely arbitrary shaped cluster is unfeasible computationally and circular shaped clusters provide a good trade-off between computational cost and meaningful and practical solutions. However, there are situations when a truly irregularly shaped cluster could be of concern such as a narrow zone along a river or an avenue. To solve this problem in the purely spatial cluster detection context, several authors have proposed spatial scan statistics using an irregular shaped scanning window (Duczmal and Assunção, 2004; Patil and Taillie, 2003; Duczmal et al., 2006; Tango and Takahashi, 2005; Assunção et al., 2006). Such scanning windows could also be adopted for our space–time statistic with some cost in additional computing time.

Our method has many desirable features. First, it does not require information about the population at risk, only cases are necessary. Second, it adjusts for purely spatial and purely temporal clustering, and it provides statistical inference for the emerging cluster detected. Third, it does not require many input parameters and the ones it does have a clear practical interpretation. This interpretation should help the user to establish reasonable values for them. We think it will be of great use in many practical applications.

Our method has been implemented in a stand-alone C++ software as well as in TERRAVIEW, a free GIS software based on the open-source TERRALIB library (see http://www.dpi.inpe.br/terraview/index.php). A suite of R functions has also been developed and they are available upon request.

References

Assunção, R., Costa, M., Tavares, A., Ferreira, S., 2006. Fast detection of arbitrarily shaped disease clusters. Statistics in Medicine 25, 723–742.

Assunção, R., Maia, A., 2007. A note on testing separability in spatial-temporal marked point processes. Biometrics 63, 290–294.

Assunção, R., Tavares, A., Correa, T., Kulldorff, M., 2007. Space–time cluster identification in point processes. The Canadian Journal of Statistics 35, 9–25.

Buckeridge, D.L., Burkom, H., Campbell, M., Hogane, W.R., Moore, A.W., 2005. Algorithms for rapid outbreak detection: A research synthesis. Journal of Biomedical Informatics 38, 99–113.

Buehler, J.W., Hopkins, R.S., Overhage, J.M., Sosin, D.M., Tong, V., CDC Working Group, 2004. Framework for evaluating public health surveillance systems for early detection of outbreaks: Recommendations from the CDC Working Group. Morbidity and Mortality Weekly Report 7, 1–11.

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

Duczmal, L.H., Assunção, R.M., 2004. A simulated annealing strategy for the detection of arbitrary shaped spatial clusters. Computational Statistics and Data Analysis 45, 269–286.

Duczmal, L.H., Kulldorff, M., Huang, L., 2006. Evaluation of the spatial scan statistics for irregular shaped clusters. Journal of Computational and Graphical Statistics 15, 428–442.

Frisén, M., 2003. Statistical surveillance. Optimality and methods. International Statistical Review 71, 403–434.

Hardy, A., 2001. Methods of outbreak investigation in the "Era of Bacteriology" 1880–1920. Social and Preventive Medicine 46, 355–360.

Henderson, D.A., 1999. The looming threat of bioterrorism. Science 283, 1279–82.

Höhle, M., 2007. Surveillance: An R package for the monitoring of infectious diseases. Computational Statistics 22, 571–582.

Höhle, M., Paul, M., 2008. Count data regression charts for the monitoring of surveillance time series. Computational Statistics and Data Analysis 52, 4357–4368.

Kenett, R.S., Pollak, M., 1996. Data-analytic aspects of the Shiryayev–Roberts control chart: Surveillance of a non-homogeneos Poisson process. Journal of Applied Statistics 23, 125–137.

Kulldorff, M., 2001. Prospective time periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society, Series A 164, 61–72.

Kulldorff, M., 2006. SaTScanTM v7.0: Software for the Spatial and Space–time Scan statistics. Information Management Services, Inc., Available at http://www.satscan.org/.

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R.M., Mostashari, F., 2005. A space time permutation scan statistic for disease outbreak detection. PLoS Medicine 2, 216–224.

Lawson, A.B., Kleinman, K., 2005. Spatial and Syndromic Surveillance for Public Health. Wiley, New York.

Marshall, J.B., Spitzner, D.J., Woodall, W.H., 2007. Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time. Statistics in Medicine 26, 1576–1593.

Mevorach, Y., Pollak, M., 1991. A small sample size comparison of the Cusum and the Shiryayev-Roberts approaches to change point detection. American Journal of Mathematical and Management Sciences 11, 277–298.

Patil, G.P., Taillie, C., 2003. Geographic and network surveillance via scan statistics for critical area detection. Statistical Science 18, 457–465.

Pollak, M., 1985. Optimal detection of a change in distribution. Annals of Statistics 13, 206–227.

Pollak, M., Siegmund, D., 1985. A diffusion process and its application to detecting a change in the drift of Brownian motion. Biometrika 72, 267–280.

Roberts, S.W., 1966. A comparison of some control chart procedures. Technometrics 8, 411–430.

Rodeiro, C.L.V., Lawson, A.B., 2006. Monitoring changes in spatio-temporal maps of disease. Biometrical Journal 48, 463–480.

Rogerson, P.A., 2001. Monitoring point patterns for the development of space–time clusters. Journal of the Royal Statistical Society, Series A 164, 87–96.

Shiryaev, A.N., 1963. On the detection of disorder in a manufacturing process. Theory of Probability and its Applications 8, 247–265.

Sonesson, C., Bock, D., 2003. A review and discussion of prospective statistical surveillance in public health. Journal of the Royal Statistical Society, Series A 166, 5–21.

Tango, T., Takahashi, K., 2005. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geography 4, www.ij-healthgeographics.com/content/4/1/11 [05 October 2005].

Waller, L.A., Gotway, C.A., 2004. Applied Spatial Statistics for Public Health Data. John Wiley & Sons, New York.

Williams, E.H., Smith, P.G., Day, N.E., Geser, A., Ellice, J., Tukei, P., 1978. Space–time clustering of Burkitt's lymphoma in the West Nile district of Uganda: 1961–1975. British Journal of Cancer 37, 109–22.

Yakir, B., 1994. Optimal detection of a change in distribution when the observations are independent, Technical Report, Department of Statistic, Hebrew University of Jerusalem.