Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 2753 – 2768
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Detecting local regions of change in high-dimensional criminal or terrorist point processes☆

Michael D. Porter*, Donald E. Brown

Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904, USA

Received 9 January 2006; received in revised form 2 July 2006; accepted 3 July 2006
Available online 25 July 2006

Abstract

A method is presented for detecting changes to the distribution of a criminal or terrorist point process between two time periods using a non-model-based approach. By treating the criminal/terrorist point process as an intelligent site selection problem, changes to the process can signify changes in the behavior or activity level of the criminals/terrorists. The locations of past events and an associated vector of geographic, environmental, and socio-economic feature values are employed in the analysis. By modeling the locations of events in each time period as a marked point process, we can then detect differences in the intensity of each component process. A modified PRIM (patient rule induction method) is implemented to partition the high-dimensional feature space, which can include mixed variables, into the most likely change regions. Monte Carlo simulations are easily and quickly generated under random relabeling to test a scan statistic for significance. By detecting local regions of change, not only can it be determined if change has occurred in the study area, but the specific spatial regions where change occurs is also identified. An example is provided of breaking and entering crimes over two-time periods to demonstrate the use of this technique for detecting local regions of change. This methodology also applies to detecting regions of differences between two types of events such as in case–control data.

Keywords: Change detection; Marked spatial point process; PRIM; Random labeling; Spatial scan statistic

1. Introduction

This paper considers the statistical detection of changes in a criminal or terrorist process. We consider any criminal or terrorist process that involves some level of intelligent site selection; that is, the criminal judiciously selects the location of the attack or crime according to their preferences or some perceived utility of that location. By observing the attack/crime locations and the value of some geographical, socio-economic, environmental, or other features of the event locations, we attempt to infer changes in the preferences or behavior of the criminals by detecting local changes in the locations and feature values of the events between two time periods.

Consider as a motivating example the locations of improvised explosive devices (IEDs) in Iraq. These deadly weapons pose a great danger to coalition forces. McFate (2005) describes the organization and increased sophistication in the strategies of deploying IEDs by the insurgency in Iraq. If we assume the locations of the events are selected in an

☆ This work is partially supported by Army Research Laboratory Grant DAAD19-03-2-0027.
* Corresponding author. Tel.: +1 434 924 5393; fax: +1 434 982 2972.
E-mail address: mdp2u@virginia.edu (M.D. Porter).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.07.002

intelligent manner, changes in the site selections could reveal behavioral changes, or those related to the terrorist decision making, such as if a new leader is planning attacks, or a security intervention has forced changes to where and how a terrorist operates. We also attempt to detect changes in the activity level of the terrorist process such as when terrorists have received additional supplies or members. We note the same type of change detection can be useful for criminal processes as well. The detection of a change could be useful for resource allocation or further investigations into the cause or details of the change.

Many spatial analyses of event data consider *only* the locations of events and leave out covariates. In considering the location data, the events are related to each other only through their geodesic distance. Of course, this type of analysis has shown its value as much can be inferred from the locations alone (Diggle, 1983). There exist some methods to detect change regions in a criminal point process, but these only consider spatially related changes (Rogerson, 2001; Rogerson and Sun, 2001; Ratcliffe, 2005). There is a similar problem in disease modeling where the interest is to detect a disease outbreak by finding regions where the disease rate varies significantly from a possibly varying baseline rate; yet these also only consider spatial changes (Besag and Newell, 1991; Kulldorff et al., 2003). However, in attempting to address the possibilities posed above, spatial displacement alone might not be sufficient, and additional data might be necessary to detect such changes in the behavioral aspects of the criminal/terrorist process. One approach could be to treat the feature data as another *spatial* dimension and use the existing methods. But that might prove difficult in this case since there is the possibility of several heterogeneous criminal decision makers (gang leaders, terrorist cells, etc.) operating in the same region. Either due to geography or preference, each decision maker might consider only a subset of the feature set and these subsets can differ between the decision makers. Additionally, these methods do not expand well to high-dimensional or mixed variable data and the feature set is not uniformly distributed in the study region as the spatial location is. Other analyses, such as time series, can consider the feature data in modeling the criminal process, but leave out the spatial aspect. Only global changes are really considered and local changes that occur only in certain spatial regions of the study area could be overlooked. Spatial regression (Anselin, 1988; Fotheringham et al., 1998) can account for the spatial regions, but requires a model-based approach that will prove difficult in the high-dimensional, mixed variable, multiple decision making scenario we are considering.

In this paper, we take an approach for explaining criminal/terrorist actions based on ecological theory. Ecology theories seek to describe the motivations and acts of crime based on the general features of one's environment, which can directly relate to the environment of the criminal or crime scene (Cohn and Felson, 1979; Brantingham and Brantingham, 1981; Byrne and Sampson, 1986; Anselin et al., 2000). In other words, the criminals/terrorists choose the location of the crime based upon some attributes or features of the location (Cornish and Clarke, 1986). Additionally, some environmental conditions will encourage or discourage crime, such as poor economic conditions or drinking establishments (Bursik, 1988). All of these theories point to the explanation that there are some features of the locations of the crimes/attacks which are important to criminal decision making. Rossmo (1999), Liu and Brown (2003), and Law and Haining (2004) use this concept to model crime by including a location's attributes or features in addition to the spatial information.

Therefore to detect changes in the criminal/terrorist process, we must detect changes not just in spatial location alone, but also in these attributes of the locations. For each type of crime or terrorist act, we can expect that the set of features that are used for decision making differ. For example, a burglar might be interested in locations that offer quick escape or hiding while a suicide bomber might choose locations that are highly populated. Indeed there might even be several groups of criminals or terrorists that commit the same crime type but have different motivations and risk levels thus consider different feature sets in their site selection.

We would like to capture these environmental factors which are influencing criminal/terrorist actions. However, since they are specific to the individual criminal and generally unknown, we must assume that not all of the features can be captured in our feature set. Therefore, guided by theory or expert opinion, we capture the features that are assumed to influence the decision making process or surrogate features that could be correlated with the true features considered. Considering the many motivations for site selection among the possibly diverse criminal and terrorist decision makers, the task of feature selection may seem daunting. This is more so in the case of change detection since a change in the process over any single feature signifies a change and one feature that was not considered in one time period could be considered in the next. We do not treat feature selection in this paper but do report on some results when excess noise variables are present in Section 4.

2. Formulation

A point process N is a stochastic model governing the location (and number) of events in some set $\mathscr{X}$ (Cressie, 1993). A point process can be represented by

$$N(\cdot) = \sum_{i=1}^{K} I\left(s_i \in \cdot\right),$$

where I is the indicator function, s_i is an event, and K is an integer-valued random variable. Therefore, $N(B)$ is the number of events in a set $B \subseteq \mathscr{X}$.

A Poisson point process satisfies two conditions (Karr, 1986):

(1) Whenever $B_1, \ldots, B_n \in \mathscr{B}$ are disjoint, the random variables $N\left(B_1\right), \ldots, N\left(B_n\right)$ are independent.
(2) For each $B \in \mathscr{B}$ and $k \geqslant 0$,

$$P\{N(B)=k\} = \frac{\exp^{-\mu(B)}\mu(B)}{k!}.$$

The σ-finite mean measure μ is such that $\mathbb{E}[N(B)]=\mu(B)=\int_B \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s} < \infty$ for any compact Borel set B. The nonnegative intensity function, when it exists, is given by $\lambda(\mathbf{s}) = \lim_{\ell(\mathrm{d}\mathbf{s})\to 0} \mathbb{E}[N(\mathrm{d}\mathbf{s})]/\ell(\mathrm{d}\mathbf{s})$, where $\ell(\cdot)$ is the Lebesgue measure. When $\lambda(\mathbf{s})$ is not a constant, but a deterministic function of $\mathbf{s}$, the process is termed nonhomogeneous (or inhomogeneous) Poisson.

Since we want to test for change between two time periods, we let the criminal process be represented by a Poisson *marked point process* defined on the product space $\mathscr{X}=A\times K$, $A \subset \mathbb{R}^2$, $K=\{1, 2\}$, where A defines a bounded geographic region of study where the events are observed and K defines the mark space representing the labels that the events receive. The labels signify the time period in which the event occurred. The events are the set $\{(\mathbf{s}_i, k_i) : \mathbf{s}_i \in A,\ k_i \in K\}$, where each event is designated by a pair of coordinates, say longitude and latitude $\left(\text{i.e. } \mathbf{s}_i = \left(s_{i,1}, s_{i,2}\right)\right)$.

The *feature space* G, defines the additional feature information relating to the locations in the study region. Define $G(B) = \{\mathbf{g}(\mathbf{s}) : \mathbf{s} \in B\}$, where $\mathbf{g}$ is the function assigning the feature values to each spatial location. Then the values of an event's location in feature space can be designated by $\mathbf{g}\left(\mathbf{s}_i\right) = \left[g_1\left(\mathbf{s}_i\right), g_2\left(\mathbf{s}_i\right), \ldots, g_p\left(\mathbf{s}_i\right)\right] \in G$ where there are p possibly mixed variables taking real (perhaps discrete), ordered, or categorical values. It is important to note that the value $\mathbf{g}(\mathbf{s})$ is assumed to be a known function of location, so given any $\{\mathbf{s}\}$, the value $\mathbf{g}(\mathbf{s})$ can be determined without error. For example, a GIS (geographical information system) can be used to extract the feature values for a set of events. A standard GIS is capable of calculating distances from the set of events to geographic objects (i.e. roads, landmarks) and determining if an event falls within a boundary (i.e. census tract, neighborhood watch area). We assume the geographically determined feature values are static within the study time horizon, that is, we are not considering any of these features in which the values will change during the period under observation. This is not much of a concern for features relating to distances which will rarely, if ever change. However, some other features (i.e. socio-economic) can change slowly over time, so care must be taken in setting the time horizon of the study if these variables are included.

We have been implying that the values of the intensity function are dependent upon the feature values. However, since only the $\mathbf{s}$ are needed to determine $\mathbf{g}(\mathbf{s})$, we will suppress the notation of $\mathbf{g}(\mathbf{s})$ in the intensity function:

$$\lambda(\mathbf{s}, k) \quad \Leftrightarrow \quad \lambda(\mathbf{s}, k; g(\mathbf{s})).$$

In a marked point process, the *ground process* $N_\gamma = \{\mathbf{s}_i\}$ represents the locations of all the events, regardless of mark, and can constitute a point process of its own. A marked point process N has *independent marks* if, given the ground process $N_\gamma = \{\mathbf{s}_i\}$, the $\{k_i\}$ are mutually independent random variables such that the distribution of k_i depends only on the corresponding location $\mathbf{s}_i$ (Daley and Vere-Jones, 2003). Assuming independent marks, the intensity function of the criminal process can be written as

$$\lambda(\mathbf{s}, k) = \lambda_\gamma(\mathbf{s}) \cdot f(k|\mathbf{s}), \quad \mathbf{s} \in A, \quad k \in \{1, 2\}, \tag{1}$$

where $\lambda_\gamma(\mathbf{s})$ is the intensity of the ground process and $f(k|\mathbf{s})$ is the mark density. We observe a realization of the marked spatial point process N in a geographical region A and record the events, over two time periods, as

$$\Omega = \left[\left(s_1, k_1\right), \left(s_2, k_2\right), \ldots, \left(s_n, k_n\right), N(A, K) = n\right]$$

and we wish to test the equivalence of the component processes N_1 and N_2 (where $N_i = N(\cdot \times \{i\})$) based on these observations. If we assume the criminal point process N has a ground process that is nonhomogeneous Poisson with independent marks, then it is completely specified by the intensity function given by (1). Thus, testing the equivalence of N_1, N_2 can be carried out by testing the equivalence of $\lambda(\mathbf{s}, 1), \lambda(\mathbf{s}, 2), \ \forall \mathbf{s} \in A$.

Note that the intensity of the spatial marked point process is actually dependent on the time period of the observations in such a way that the spatial intensity is nondecreasing in the length of observation time. Since we wish to compare intensities over two time periods which might not be of equal duration this necessitates a correction to each component intensity. Let $H(\tau_i) = \int_A \lambda(\mathbf{s}, k=i)\, d\mathbf{s}$ be the measure of the temporal effects in the time period τ_i, which could include temporal trends. Thus, the following formulation will allow testing with unequal time periods or against some known or assumed time trends in the intensity under the null hypothesis of no change

$$\begin{aligned} \mathrm{H}_0 : \lambda(\mathbf{s}, k=1)/H(\tau_1) &= \lambda(\mathbf{s}, k=2)/H(\tau_2) \\ &\Rightarrow \lambda_\gamma(\mathbf{s}) \cdot f(k=1 \mid \mathbf{s})/H(\tau_1) = \lambda_\gamma(\mathbf{s}) \cdot f(k=2 \mid \mathbf{s})/H(\tau_2) \\ &\Rightarrow \frac{f(k=1 \mid \mathbf{s})}{f(k=2 \mid \mathbf{s})} = \theta(\mathbf{s}) = \frac{H(\tau_1)}{H(\tau_2)}. \end{aligned}$$

The parameter of interest, $\theta(\mathbf{s})$, is only dependent on the ratio of the mark densities. In the case of equal time periods (and no time trends), $\theta(\mathbf{s}) = 1$ under the null. Under the alternative hypothesis, there exists a region $B \subseteq A$ where the ratio of mark densities deviates from the null

$$\begin{aligned} \mathrm{H}_0 : \theta(\mathbf{s}) &= \frac{H(\tau_1)}{H(\tau_2)} \quad \forall s \in A, \\ \mathrm{H}_\mathrm{a} : \theta(\mathbf{s}) &= \begin{cases} \delta \neq \dfrac{H(\tau_1)}{H(\tau_2)} & \forall\, s \in B \subseteq A, \\ \dfrac{H(\tau_1)}{H(\tau_2)} & \forall\, s \in A \backslash B. \end{cases} \end{aligned} \tag{2}$$

The null hypothesis of no change in the intensity parameter anywhere in the study region A is tested against the alternative of a change in intensity only in a subset $B \subseteq A$, by testing the parameter δ within a region B.

After posing the problem in the form of the hypothesis test specified above, what remains is to (1) select an appropriate test statistic to test the hypothesis, (2) specify how the region B for testing will be identified, (3) establish significance testing and (4) evaluate the results. The first three will be discussed in Section 3. Section 4 details results from a simulation study and Section 5 provides an example of applying this methodology to a data set involving breaking and entering crimes in Richmond, VA.

3. Methodology

This section describes the methodology we take for detecting the local regions of change in the criminal or terrorist point process. The method consists of selection of a test statistic, finding the possible regions where change could have occurred, and testing for significance of the possible change regions. We select a generalized likelihood ratio statistic for hypothesis testing. This statistic reduces to a function of the number of events of each type in region B. A modified PRIM methodology (see below) is used to search for the most likely change regions. Finally, we use a Monte Carlo approach via random labeling to estimate the p-value of the observed test statistic.

3.1. Likelihood ratio test

The generalized likelihood ratio test (GLRT) is used for testing the hypothesis given by (2). The GLRT is a composite hypothesis test which has as its statistic the ratio of the likelihood of the observations under the two hypotheses where the parameters are estimated by their maximum likelihood values. Let T be the GLRT statistic for a potential change region $B \subseteq A$

$$T(B) = \frac{\mathscr{L}\left(\theta_{\mathrm{H}_0}\right)}{\sup_{\theta \in \Theta} \mathscr{L}(\theta, B)} = \frac{\mathscr{L}\left(\theta_{\mathrm{H}_0}\right)}{\mathscr{L}\left(\hat{\theta}_{\mathrm{ML}}, B\right)}, \tag{3}$$

where $\theta = [\theta(s \in B), \theta(s \in A \backslash B)]$ and $\Theta \in [0, \infty]^2$. Therefore, $\theta_{H_0} = [H(\tau_1)/H(\tau_2), H(\tau_1)/H(\tau_2)]$ and $\hat{\theta}_{ML} = \left[\hat{\delta}_{ML}, H(\tau_1)/H(\tau_2)\right]$. Under the null, $\theta(\mathbf{s})$ is specified uniquely, hence H_0 is simple. However, H_a is composite with $\hat{\theta}$ including the maximum likelihood estimate of $\delta \in [0, \infty]$ in region B. $T(B)$ takes values in [0, 1]. Under H_0, $T(B)$ will be closer to 1, so H_0 will be rejected when $T(B)$ is small (see Section 3.3).

Assuming independent marks, the GLRT statistic, given a realization Ω is

$$T(B) = \prod_{i=1}^{N_{1,2}} \left[\frac{f\left(k_i \,\middle|\, \mathbf{s}_i; \theta_{H_0}\right)}{f\left(k_i \,\middle|\, \mathbf{s}_i; \hat{\theta}_{ML}\right)} \right], \tag{4}$$

where $N_{1,2}$ is the number of events of either mark in A. Since $f(k=1 \mid \mathbf{s}) + f(k=2 \mid \mathbf{s}) = 1$ for all $\mathbf{s}$, this implies

$$\begin{aligned} f(k=1 \mid \mathbf{s}) &= \theta(\mathbf{s})(\theta(\mathbf{s})+1)^{-1}, \\ f(k=2 \mid \mathbf{s}) &= (\theta(\mathbf{s})+1)^{-1}. \end{aligned} \tag{5}$$

Rewriting (4) in terms of (5) leads to

$$\begin{aligned} T(B) &= \prod_{i=1}^{N_{1,2}(B)} \left[\frac{\theta_{H_0}\left(\theta_{H_0}+1\right)^{-1}}{\hat{\theta}_{ML}\left(\hat{\theta}_{ML}+1\right)^{-1}} \right]^{y_i} \left[\frac{\left(\theta_{H_0}+1\right)^{-1}}{\left(\hat{\theta}_{ML}+1\right)^{-1}} \right]^{(1-y_i)} \\ &= \left[\theta_{H_0}\right]^{N_1} \left[\frac{1}{\left(\theta_{H_0}+1\right)} \right]^{N_{1,2}(B)} \left[\frac{N_{1,2}(B)}{N_1(B)} \right]^{N_1(B)} \left[\frac{N_{1,2}(B)}{N_2(B)} \right]^{N_2(B)} \\ &= \left[\frac{\theta_{H_0} \cdot N_{1,2}(B)}{\left(\theta_{H_0}+1\right) \cdot N_1(B)} \right]^{N_1(B)} \left[\frac{N_{1,2}(B)}{\left(\theta_{H_0}+1\right) \cdot N_2(B)} \right]^{N_2(B)}, \end{aligned} \tag{6}$$

where $y_i = 1$ if $k_i = 1$, $N_i(B) =$ Number of i-labeled events in region B, and $\hat{\theta}_{ML} = N_1(B)/N_2(B)$. As seen from (6), T is only dependent on the number of events in region B; there is no dependence on the size of B.

3.2. Finding region B^*

The change region B defines the geographical subset of A where change has occurred in the intensity function of the point process. For testing the hypothesis in (2), we must identify this region B. To do this, we could conduct our search over the spatial region A. However, as previously mentioned, we want to be able to detect the changes in feature space, because we assume that it is the values of these attributes that are influencing the actors' site selection process. Therefore, we do not search in geographic space only, but include feature space (i.e. we search over $A \times G$) for the region B^* that provides the minimum value of $T(B)$, thus providing the most likely candidate region for rejection of the null. This procedure of searching for extrema regions for significance testing is termed *scan process* (Priebe et al., 1997, 2001) or boundary crossing problem (Loader, 1991).

3.2.1. Scan process

In general, a *scan process* is used to detect significant clusters of events; that is, detecting local regions where the number of events is elevated far beyond expectation, then testing the significance of the deviation while accounting for multiple hypothesis testing (since we searched over all possible regions) (Glaz and Balakrishnan, 1999). The scan process consists of creating a window, W_x, of some geometry (usually a convex set), moving this window over the entire region of interest ($\forall x \in \mathscr{X}$), and calculating some score for each window, $S(W_x)$. The multiple hypothesis problem is resolved by only testing on the *scan statistic*, the maximum score over all windows searched. The scan statistic, $SS = \sup_{x \in \mathscr{X}} S(W_x)$, is dependent on the geometry selected for W. However, multiple window geometries (size and shape) can be considered to allow detection when the geometry of the change region is unknown a priori (Loader, 1991).

We adopt the same process (with a few adjustments) to detect significant regions of change instead of clusters. Since we want to find the *minimum* value, the scan statistic becomes the value of our GLR statistic at region B^*

$$SS(\Omega) = \inf_{B \in \mathscr{B}} T(B) \\ = T\left(B^*\right), \tag{7}$$

where $\mathscr{B}$ is the set of windows that we search over.

Since we are trying to detect change in the criminal or terrorist behavior process, we do not set these windows in geographic space only, but include feature space (i.e. $\mathscr{X} = A \times G$). Additionally, since we do not know the possible geometries of the change regions, we want inclusion of a broad number of geometries for calculating the scan statistic. However, we generally assume in the alternative that there is only one change region so we will restrict the possible change regions $\mathscr{B}$ to be connected sets in $\mathscr{X}$, but without restriction on the size of the regions. Furthermore, we will ease computational complexity by assuming the regions are also hyper-rectangles in the real variables. Since T only depends on the number of events and not the size of the region, there is a finite search space when the number of events in the study region is finite. However, when the number of events and features is high, searching over all possible combinations is computationally expensive. We proceed by searching in a smart way to find good approximations to the scan statistic. That is, we only search over a restricted set $\mathscr{B}' \subseteq \mathscr{B}$ and hope this set includes B^*, or a large portion of it.

By including feature space in $\mathscr{X}$, the potential change regions are the hyper-rectangles in feature space, geographic space, or a combination of the two. A hyper-rectangle in feature space does not translate into a hyper-rectangle (or even a connected set) when projected into geographic space. For example, consider as a possible change region the locations less than distance d_1 from a road and in a census tract with population greater than d_2. While this is a rectangle in feature space, the same region will appear as a line segment, with width d_1, following the road networks only in census tracts that are valued over d_2 when viewed in geographic (or map) space. Thus, the identified change regions will not take any specific shape in geographic space and a single change region (that is a connected set in $\mathscr{X}$) might appear as several disjoint sets when projected onto geographic space.

3.2.2. *PRIM*

We choose to use a modification of PRIM (patient rule induction method) as an expedient way to search for B^*, thus the value of the scan statistic (Friedman and Fisher, 1999). PRIM handles high dimensional and mixed data well and can easily be adapted to search for low values of $T(B)$. PRIM finds the *boxes* (hyper-rectangular subregions for real-valued variables) where the response (e.g. T) is low (or high).

We start with a box $B_1 = A \times G$ that covers the entire study region. Our method proceeds by producing a series of boxes, $\{B_c\}$, decreasing in size, by successively peeling away a subbox in such a manner that each new box B_{c+1} has the lowest value for $T(B_{c+1})$ among all possible subboxes. Let the box B_c contain $N(B_c)$ events and $\{\mathbf{s}_i \in \Omega\}$ be all the events in both time periods. The candidate subboxes for peeling at each step are

$$b^-_{j,c}(\alpha_c) = \left\{\mathbf{s}_i \in \Omega : g_j(\mathbf{s}_i) \leqslant g_j(\alpha_c)\right\},$$

$$b^+_{j,c}(\alpha_c) = \left\{\mathbf{s}_i \in \Omega : g_j(\mathbf{s}_i) \geqslant g_j(1-\alpha_c)\right\},$$

$$b^m_{j,c} = \left\{\mathbf{s}_i \in \Omega : g_j(\mathbf{s}_i) = s_j(m)\right\} \quad \left(\text{if } g_j \text{ is categorical}\right)$$

for $j = 1, 2, \ldots, p+2$, with the construction $g_i(\mathbf{s}) = \mathbf{s}_i$, $i = (1, 2)$ for the two spatial coordinates and $i = 3, 4, \ldots, p+2$ for the p feature variables. The values $g_j(\alpha_c)$ and $g_j(1-\alpha_c)$ are the α_c-quantile and $(1-\alpha_c)$-quantile of the g_j values for all the events within the current box, B_c, respectively. The $\alpha_c \in (0, 1)$ dictates the size of the subboxes under consideration for peeling at step c. There are $p+2$ variables since we include the two spatial dimensions with the p feature variables. When g_j is a categorical variable, the set of subboxes $\left\{b^m_{j,c}\right\}_{m=1}^{M_j}$ is available for peeling where M_j indexes the categories still included in box B_c.

The optimal subbox for peeling at step c becomes

$$b^*_c = \mathop{\arg\max}_{\substack{j \in \{1,2,\ldots,p+2\} \\ \phi \in \{+,-,m\}}} T\left(b^{\phi}_{j,c}\right)$$

for a given α_c. This creates the new box $B_{c+1} = B_c \backslash b_c^*$. Peeling away the set with the largest T is equivalent to retaining the new box with the smallest value of T.

We continue peeling until T can no longer be decreased beyond the minimum at step c. Since T is only dependent on the number of events of each type, we can find the minimum possible value of T (from any given box) if we continued peeling. It can be seen from (6) that the minimum value of T will occur when all the events have the same label. The stopping rule becomes

if $\min\left(\left[\dfrac{\theta_{\mathrm{H}_0}}{(\theta_{\mathrm{H}_0}+1)}\right]^{N_1(B_c)}, \left[\dfrac{1}{(\theta_{\mathrm{H}_0}+1)}\right]^{N_2(B_c)}\right) \leqslant \min\limits_{1,2,\ldots,c} T(B_i)$ **then**
 continue peeling
else
 $SS' = \min\limits_i T(B_i)$
$B^{*'} = B_{j^*}$ where $j^* = \arg\min\limits_j T(B_j)$
end if

The result of this peeling step is a single box in space $\mathscr{X}$. After the peeling procedure, it is possible that the best box can be made better with a pasting step. Additional events can be added (pasting) in a similar manner to peeling by adding a subbox to $B^{*'}$. That is, the final box becomes $\widetilde{B^*} = B^{*'} \cup b_P^*$, where b_P^* is the subbox that minimizes T out of all eligible subboxes for the pasting step and $\widetilde{SS} = T\left(\widetilde{B^*}\right)$. The class of eligible subboxes is defined as those boxes, of any size, that extend the current box on one dimension (variable) but maintain the current box boundaries on all other variables. For a categorical variable g_j, the eligible subboxes are those created by adding the values not included in $B^{*'}$.

One could imagine any number of modifications to this procedure such as incorporating pasting into the peeling steps, or peeling nonorthogonally. The main concept is to estimate the true B^* in a computationally simpler way than searching combinatorially.

Although the patient rule helps to find the global optima, it is not guaranteed to find it. In fact different choices of initial peels can lead to different final boxes. The peeling fraction α_c is the parameter impacting the box succession. The standard PRIM uses a constant, usually $0.05 \leqslant \alpha \leqslant 0.10$ (Friedman and Fisher, 1999). Since we are searching for the global minimum we want to avoid entrapment in a local minimum. Therefore, we let α_c be a variable and run PRIM for J iterations each time modifying the values for the variable α_c's. This will create J values for $\widetilde{B^*}$ and $\widetilde{SS}$ and we hope one of these values is the global minimum. The final estimates lead to a test of significance of $\widehat{SS} = \min_j \widetilde{SS}_j$ for change in region $\widehat{B^*} = \widetilde{B^*_{j^*}}$, where $j^* = \arg\min_j \widetilde{SS}_j$.

This method aims to detect the one region where change has most likely occurred. But what if there are multiple disjoint regions of change? First, note that the value $\widehat{SS}$ only depends on the events inside region $\widehat{B^*}$, so the presence of multiple change regions in the area under study has no effect on identifying the top change region or on inference (see below). One could consider removing the events in $\widehat{B^*}$ and repeating PRIM to identify the secondary regions of possible change. However, the regions would not necessarily be hyper-rectangles (in $\mathscr{X}$) and the test for change will not be as powerful.

Since we wish to capture as many features as possible that influence the criminal or terrorist process we risk choosing some features that are only noise. The PRIM shows some resistance to the disagreeable effects of including noise variables. The final box may consist of some variables that were not used in the peeling process. This means the same results would be obtained if those variables were not included in the initial evaluation. Thus, the nature of the peeling process automatically excludes most of the noise variables from consideration.

3.3. Significance testing

Given a realization of the process, after finding the region $\widehat{B^*}$ and calculating the observed test statistic $t^{\mathrm{obs}} = T\left(\widehat{B^*}\right)$, we evaluate the evidence we have for rejecting H_0 in the form of a p-value. The p-value of an observation is the probability that under the null distribution of T you would get an observation as extreme (in this case as small) as what was observed

$$p_{\mathrm{obs}} = \Pr\left(T \leqslant t^{\mathrm{obs}}; \mathrm{H}_0\right),$$

where H_0 will be rejected if $p_{\text{obs}} \leqslant \alpha$ for some significance level $\alpha \in [0, 1]$. Small p-values will occur when t^{obs} is small.

Since the distribution of T under H_0 is not known, a Monte Carlo test is used for significance testing. A Monte Carlo test consists of generating $M - 1$ observations under H_0, calculating the values of t^m for each simulated observation, $m = 1, 2, \ldots, M - 1$, then estimating the p-value by the proportion of times $t^m \leqslant t^{\text{obs}}$.

We could generate the Monte Carlo observations by first estimating the intensities $\lambda(\mathbf{s}, k = i),\ i = (1, 2)$ then generating new event locations from this distribution (Kornak et al., 2003). However, intensity estimation will be very complicated in the high-dimensional feature space we consider. Additionally, estimation of the intensity will introduce an additional source of error. Fortunately, from (2), θ_{H_0} reduces to only a function of the mark densities, and is constant over the entire study region, A. So we can produce new realizations of the point process by *random labeling*, that is assigning for each event in the original realization Ω^{obs}, a new mark according to (5).

The procedure is to generate realizations of the marked spatial point process by random labeling the observed event marks under the null. Next, the values of the scan statistics, $\{(t^m; H_0)\}_{m=1}^{M-1}$, are calculated for each of these realizations from PRIM. Order these observations, $\left\{t^{\text{obs}}, t^m\right\}$, from smallest to largest and let l_n be the order of the nth observation. The estimated p-value then becomes

$$\hat{p}_{\text{obs}} = l_{\text{obs}}/M. \tag{8}$$

This testing scenario adjusts for multiple hypotheses by testing on the value of the scan statistic. By reformulating the change model so that the intensities are not being directly tested, we can generate realizations of the null very simply and quickly by random labeling. This procedure provides a means to detect significant local changes in high-dimensional point processes.

4. Simulated data

To evaluate the proposed methodology, we provide an example with simulated data. The estimated ROC and four other measures are examined. We also evaluate the effects of the PRIM pasting step and random noise variables.

We examine a region A that is a unit square on $[0, 1]^2$. Six features are identified for this region, in addition to the spatial coordinates, so $\mathscr{X} = A \times G$, where $G = [g_1, g_2, \ldots, g_6]$. The features $\{g_1, g_2, g_3, g_6\}$ represent distances to certain landmarks (e.g. $g_1(\mathbf{s}) =$ distance from location $\mathbf{s}$ to the nearest road). Features $\{g_4, g_5\}$ represent values of some variable recorded at the census tract level. Fig. 1 displays the values of the features in region A.

Next, we assume there are a group of insurgents operating in the region that consider three of the features of a location when planning attacks. Specifically, we assume the insurgents attack according to a nonhomogeneous Poisson process and their preferences dictate an intensity function given by

$$\begin{aligned} &\lambda_1(\mathbf{s}) = \exp\left(-c_y s_2 - c_1 g_1(\mathbf{s}) + c_5 g_5(\mathbf{s})\right) \cdot C_1, \\ &\lambda_r = 15, \end{aligned} \tag{9}$$

where $c_y = 0.5,\ c_1 = 100,\ c_5 = 1,\ C_1 = 645.2$. We assume some random attacks in the form of a homogeneous Poisson process on A with rate λ_r that we associate with group 1. The expected number of group 1 events in A becomes

$$E\left[N_1(A)\right] = \mu_1(A) = \int_A (\lambda_1(\mathbf{s}) + \lambda_r)\ \mathrm{d}\mathbf{s} = 115.$$

Then we assume an additional insurgency group becomes active in time period τ_2. This second group has different preferences for locations of attacks and operates independently of the first group. The second group's intensity function is given by

$$\lambda_2(\mathbf{s}) = I\ (g_2(\mathbf{s}) \leqslant 0.10, g_3(\mathbf{s}) \leqslant 0.05, g_4(\mathbf{s}) \leqslant 0.30) \cdot C_2, \tag{10}$$

where I is the indicator function and $C_2 = 833.3$. Fig. 2 shows the intensities of both groups mapped onto A.

Therefore, the region where change occurs is $B^* = \{\mathbf{s} : g_2(\mathbf{s}) \leqslant 0.10, g_3(\mathbf{s}) \leqslant 0.05, g_4(\mathbf{s}) \leqslant 0.30\}$ with a Lebesgue measure of $\ell\,(B^*) = 0.0424$ (i.e. B^* covers 4.24% of A). The expected number of insurgent attacks from group 2 is $\mu_2(A) = 35$. While the change region is a connected set in feature space, when projected onto A it can be seen that the region of change consists of several disjoint geographic sets.

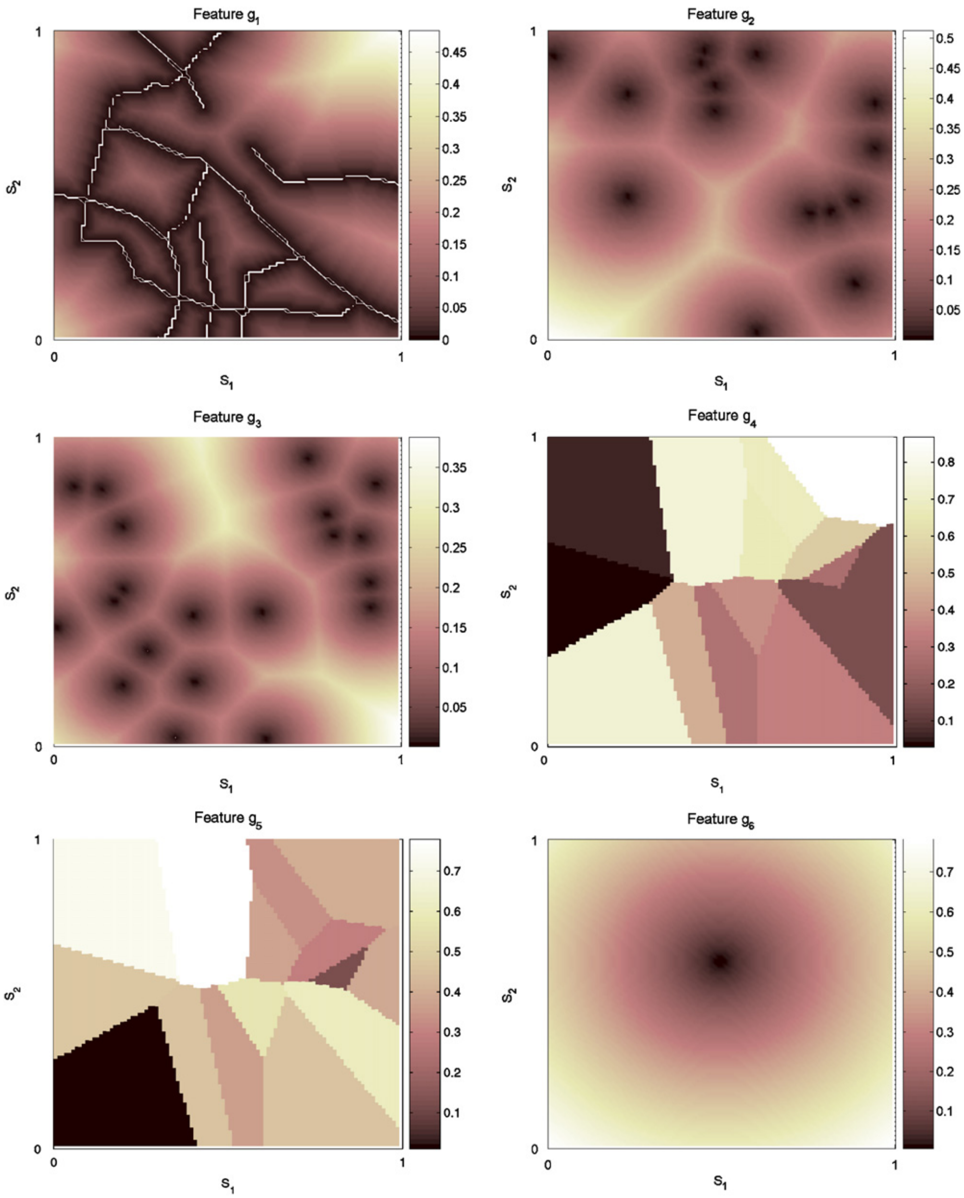

Fig. 1. Feature values.

Thus, we have $\lambda(\mathbf{s}, k=1)=\lambda_1(\mathbf{s})+\lambda_r$ for τ_1 and $\lambda(\mathbf{s}, k=2)=\lambda_1(\mathbf{s})+\lambda_r+\lambda_2(\mathbf{s})$ for τ_2. Let the length of $\tau_1=\tau_2$ and assuming no temporal trends, according to the null hypothesis (2), $\theta(\mathbf{s})=1$ everywhere in A. To evaluate our methodology, we generated 100 realizations of the point process specified by Eqs. (9) and (10) (Møller and Waagepetersen, 2004). For each of these realizations we applied PRIM with $\alpha_c \in [0.05, 0.15]$. For each original observation, $J\,(=500)$ iterations were performed creating $\left\{\widetilde{SS}_j, \widetilde{B^*}_j\right\}$. And the minimum value of the scan statistic from the 500 iterations is retained, $\widehat{SS}=\min_j \widetilde{SS}_j$ along with $\widehat{B^*}=\arg\min_j \widetilde{B}^*_j$.

Next for significance testing, 99 simulations of the process under H_0 are created. For each simulation, the labels for the events are randomly assigned to τ_i, $(i=1,2)$ with $f(k=i|\mathbf{s})=0.50$, and $J\,(=500)$ iterations of PRIM are performed on each of these Monte Carlo observations. Then the estimates for the p-value are calculated from (8).

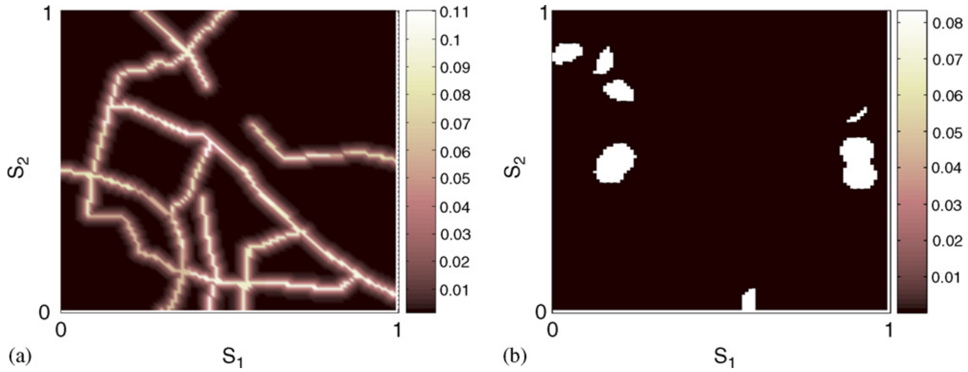

Fig. 2. Intensity function of (a) group 1, (b) group 2.

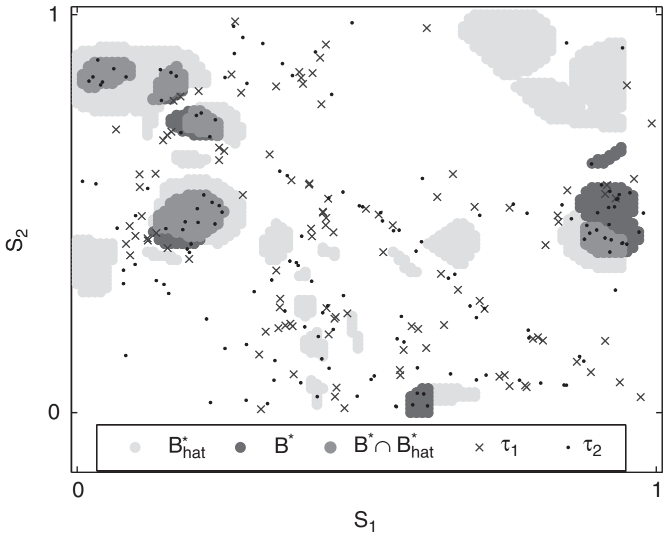

Fig. 3. Example detection region projected onto A. The change region is specified by the peels listed in Table 1. This realization resulted in $R_1 = 0.75,\ \ R_2 = 0.54,\ \ R_3 = 0.16,\ \ R_4 = 0.83$.

This results in a set $\left\{ \widetilde{B_i^*}, \hat{p}_i \right\}_{i=1}^{100}$ from which we can evaluate the methodology. This is performed under three scenarios to compare the results obtained when using the pasting step, without the pasting, and by adding four $U[0, 1]$ random noise variables.

Let E_i $(i = 1, 2)$ be the events from insurgency group i, with the random events assigned to group 1. We consider $\hat{p}$ and four other measures to evaluate the success of our method in capturing the true change region B^*, or the events, E_2 that constitute the change:

$R_1 = \left| \left\{ E_2 \cap \widehat{B^*} \right\} \right| / |\{E_2\}|$—the proportion of group 2 events captured in $\widehat{B^*}$,

$R_2 = \ell \left(\left\{ B^* \cap \widehat{B^*} \right\} \right) / \ell \left(B^* \right)$—the proportion of region B^* captured by $\widehat{B^*}$,

$R_3 = \left| \left\{ E_1 \cap \widehat{B^*} \right\} \right| / \left| \left\{ E_{1,2} \cap \widehat{B^*} \right\} \right|$—the proportion of the events captured in $\widehat{B^*}$ that are from group 1,

$R_4 = \ell \left(\left\{ \widehat{B^*} \backslash B^* \right\} \right) / \ell \left(\widehat{B^*} \right)$—the proportion of region $\widehat{B^*}$ that does not belong to B^*,

where $| \cdot |$ is the cardinality of a set and $\ell(\cdot)$ is the Lebesgue measure.

We are essentially considering two aspects of the change, the *region* where change occurs (R_2, R_4) and the *events* (R_1, R_3) that constitute the change. Since PRIM is operating on the events only, it is assumed this method would do better at capturing the events from group 2 than the region where they operate.

Table 1
Example change region defined in $\mathscr{X}$

Change region $\widehat{B^*}$	
$\mathbf{s}_1$	$0 \leqslant \mathbf{s}_1 \leqslant 0.9489$
$\mathbf{s}_2$	$0 \leqslant \mathbf{s}_2 \leqslant 1$
g_1	$0.231 \leqslant g_1 \leqslant 1$
g_2	$0 \leqslant g_2 \leqslant 0.2702$
g_3	$0 \leqslant g_3 \leqslant 0.0856$
g_4	$0 \leqslant g_4 \leqslant 1$
g_5	$0.3932 \leqslant g_5 \leqslant 0.7791$
g_6	$0.1988 \leqslant g_6 \leqslant 0.5900$

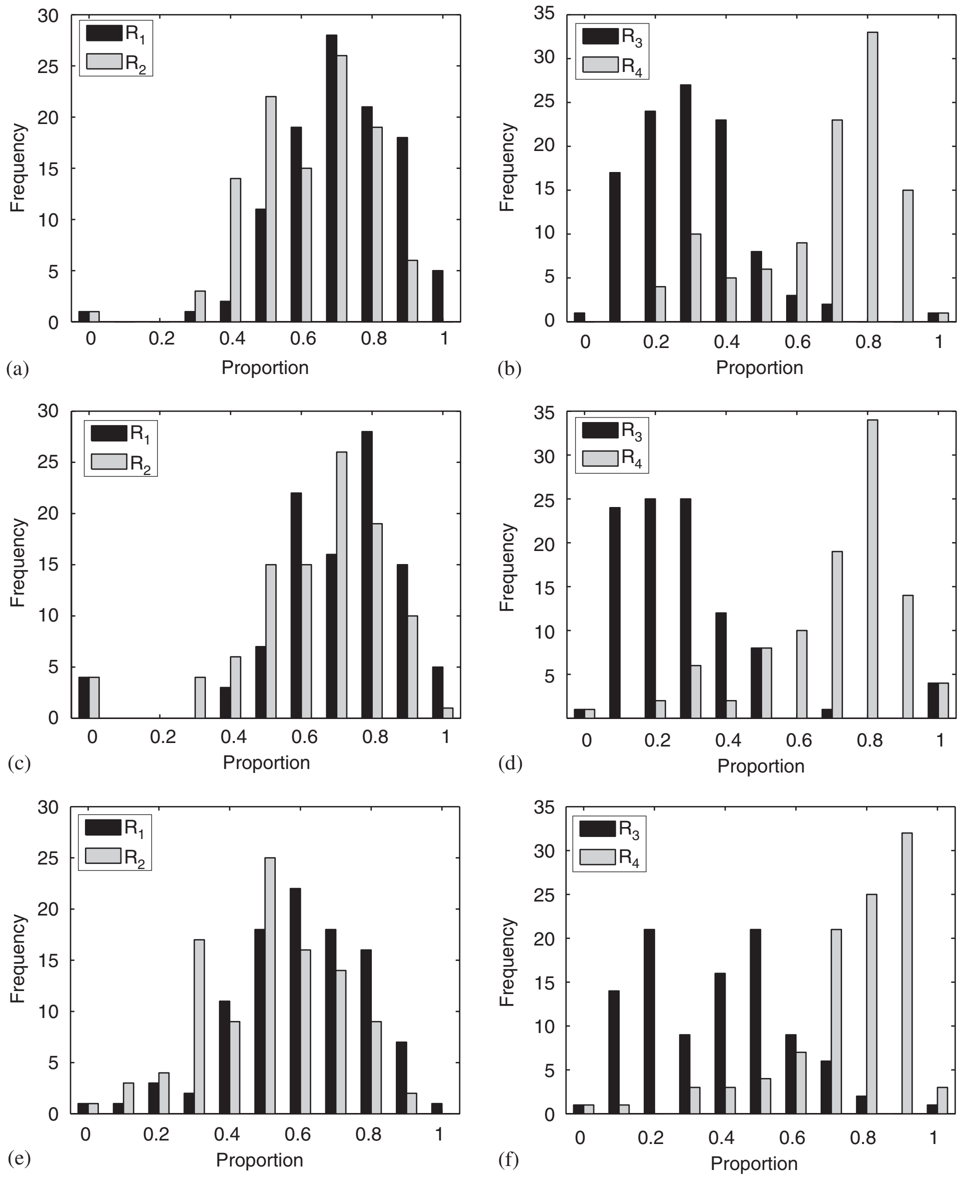

Fig. 4. Histogram of R_1, R_2, R_3, R_4 under pasting, no pasting, and added noise: (a) R_1 and R_2 under pasting; (b) R_3 and R_4 under pasting; (c) R_1 and R_2 under no pasting; (d) R_3 and R_4 under no pasting; (e) R_1 and R_2 with noise; (f) R_3 and R_4 with noise.

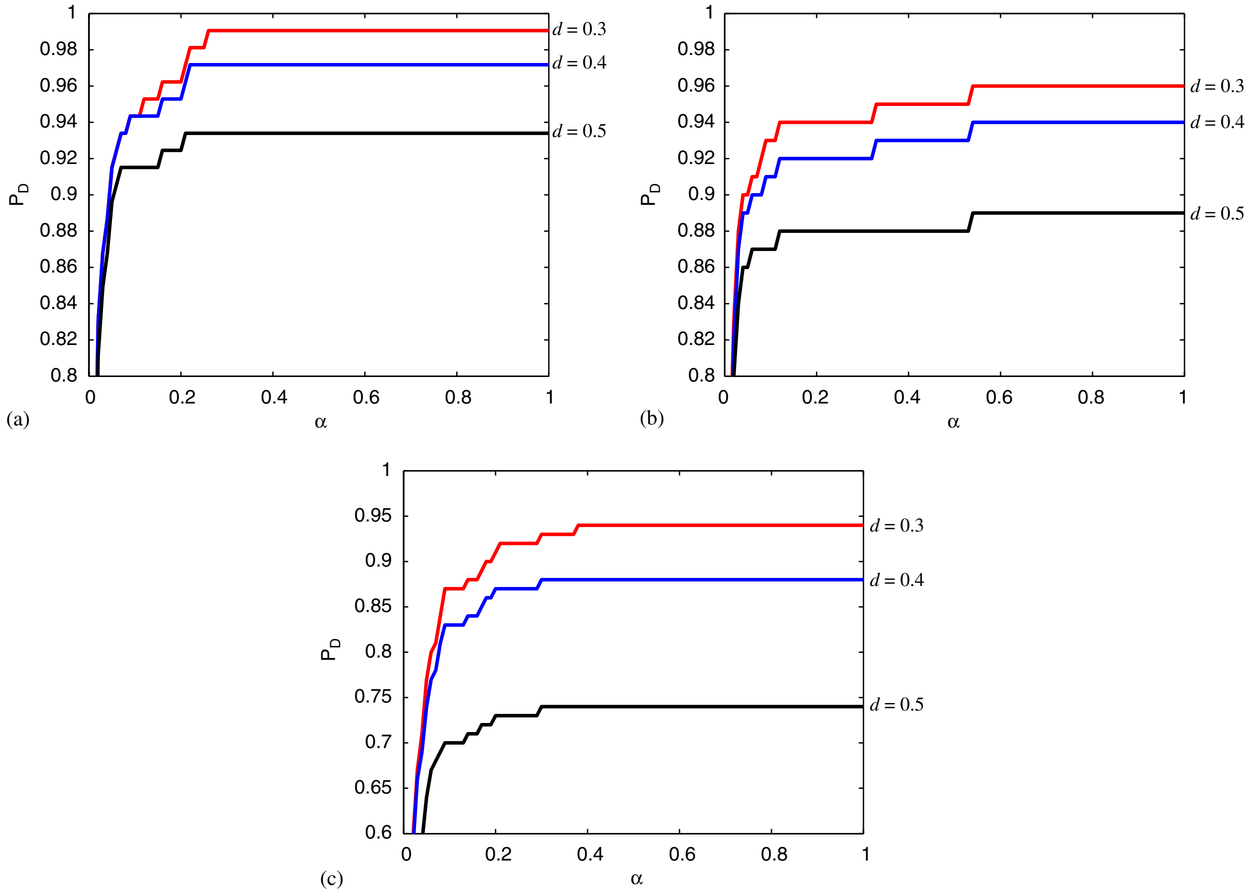

Fig. 5. ROC curves based on R_1: (a) ROC for Case 1: pasting; (b) ROC for Case 2: no pasting; (c) ROC for Case 3: noise.

Fig. 3 shows a typical detected change region from the simulation in geographic space, A and Table 1 shows the same change region in $\mathscr{X}$. The detected change region, $\widehat{B^*}$, contained 42 events marked τ_2 and only one marked τ_1 and an estimated $\hat{p}$ of 0.01. While both the true change region and the estimated change region are connected hyper-rectangles in $\mathscr{X}$, they are irregularly shaped disjoint sets when projected onto geographic space. Notice that $\widehat{B^*}$ covers a much larger area that the true change region, including regions where no events occur. This is the result of searching for change regions in feature space as opposed to only geographic space.

Fig. 4 displays the histograms for the values of R_1 to R_4 for the three scenarios: pasting, no pasting, and added noise. Within each scenario, the values of R_1 and R_2 are similar to each other, showing that PRIM identifies about the same proportion of new events as the change region. However, the values of R_4 are much larger than R_3 meaning that the PRIM method identifies the events that constitute change accurately, but estimates a change region that encompasses an area larger than the actual change region. It is also apparent that there is not much difference between the pasting and nonpasting settings. While the values for R_1 and R_2 decrease without pasting, R_3 and R_4 increase. However, both increase and decrease appear minimal. It is also apparent the addition of the noise decreases the values of R_1 and R_2 and increase the values of R_3 and R_4. This will hinder the ability of the method to detect significant change. We will evaluate this further in the following.

We next consider the receiver operating characteristics (ROC) for probability to detect (P_{D}) a significant region and α, the significance level. A detection is declared if $\hat{p} \leqslant \alpha$ and either $R_1 \geqslant d$ or $R_2 \geqslant d$ where $d \in [0, 1]$. The restriction on R_1 and R_2 ensures that the region identified includes some part of the true change region. Fig. 5 shows the ROC curves for $d = \{0.3, 0.4, 0.5\}$ under each of the three scenarios.

The ROC evaluates two aspects of the methodology: the ability of PRIM to identify B^* and the ability of T and random labeling to detect departures from the null. Notice from Fig. 5 that P_{D} does not go to 1 as $\alpha \to 1$. This is because there will be some occasions where R_1 or $R_2 < d$. In these cases, PRIM fails to detect enough of the change region. Additionally, even if PRIM identifies a good portion of the change events or change region, the use of the T statistic and random labeling might not provide a low estimate of $\hat{p}$.

This shows that, for this synthetic model, PRIM fails to specify a large percentage of the actual change region or change events about 1% of the time with pasting, 4% of the time under no pasting, and 5% with noise when $d = 0.30$. If the intensity of the change is great enough, capturing only a small portion of the overall change region will still result in a declaration of a significant change region. Additionally, this method does best at capturing the events that constitute the change, but seems to overestimate the region where change occurs.

If the change does not occur until some time into τ_2, or if μ_2 is not large, there might be no significant regions of change found. However, by providing $\hat{p}$ and $\widehat{B^*}$ this method can also be used in an exploratory nature by identifying the potential regions or events of change that will lead to further investigations.

5. Results

In this section we provide results from a criminology data set. We examine the location of breaking and entering crimes in Richmond, VA. Here we seek to discover if there is a different pattern between the crimes that occur during the summer and the rest of the year.

5.1. Breaking and entering crime locations

We evaluated residential breaking and entering (B&E) crimes reported in Richmond, VA in 1997. There has been some evidence that there is an elevation in property crimes during the summer months in accordance to the routine activities (RA) theory of crime (Cohn and Felson, 1979; Cohn and Rotton, 2000; Hipp et al., 2004). The RA theory suggests that the conditions for crime include a motivated offender, a suitable target, and an absence of guardian against crime. People tend to spend more time out of the home due to vacations or activities and more windows are left down in the summer months. The increase in B&E crimes can then be explained by an increased opportunity for the criminal to find such properties left without supervision and easily accessible. Additionally, juveniles will be out of school during this time increasing the number of possible offenders. We provide an illustrative example of how our methodology can be used to examine hypotheses related to the possibility that a different or additional group of criminals operates during these summertime conditions.

The residential B&E crimes for 1997 were partitioned into summer crimes (1 June–15 September) and nonsummer crimes (1 January–31 May, 16 September–31 December). With 107 days in summer and 258 days in the rest of the year, we must adjust for the inequality in length of the time periods. Assuming no temporal trends, the ratio $H(\tau_1)/H(\tau_2)$ is proportional to the ratio of the lengths of the time periods (=107/258 days). Thus under the null, $\theta(\mathbf{s}) = 0.4147\ \forall \mathbf{s} \in A$.

To test for change regions, the analysis considered 27 spatially specified features (including proximity measures and census values) along with the two spatial coordinates (see Table 2). All feature values were obtained from a GIS with the 1997 estimates of the census features taken from Census_CD+Maps (1998). We applied PRIM with $\alpha_c \in [0.05, 0.15]$. The procedure was performed 500 additional times, creating $\left\{\widetilde{SS}_j\right\}$. And the minimum value of the scan statistic from the 500 iterations is retained, $\widehat{SS} = \min_j \widetilde{SS}_j$. For significance testing, 999 simulations of the process under H_0 are created.

Table 2
Proximity and socio-economic features used for B&E analysis

Proximity features			
Dist_Intersection	Dist_RrhaOwned	Dist_TrafficLight	Dist_Ramp
Dist_BusStop	Dist_PoliceSta	Dist_TrPark	Dist_PlWorship
Dist_Shopping	Dist_Offices	Dist_School	Dist_Mall
Dist_Industrial	Dist_HistLand	Dist_CommunityCent	Dist_Bridge
Socio-economic features			
Tot_Pop	%Non_White	%Male_under_40	%Non_Native
%Working	Median_HH_Income	Total_HH	%Vacant_HH
%Rentals	Avg_Pop_HH	Med_Value_HU	

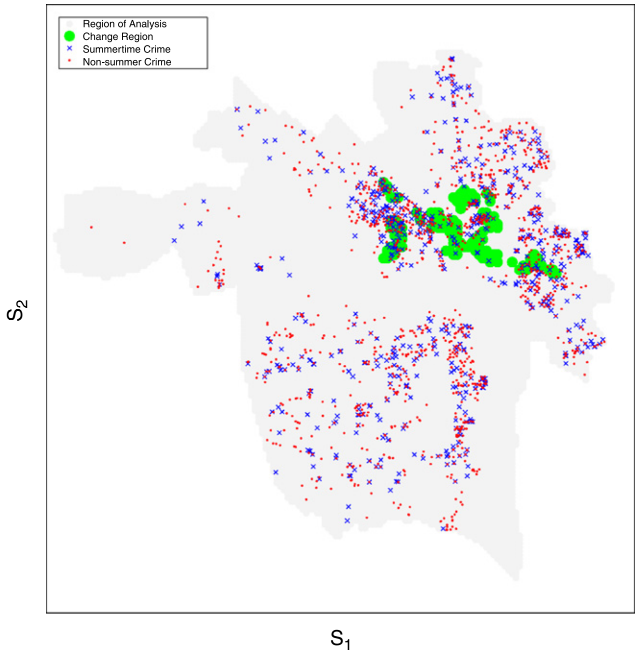

Fig. 6. Significant change location in spatial view (shaded region). When projected onto geographic space, the connected set in feature space becomes disjoint.

Table 3
Change region in $\mathscr{X}$

Change region $\widehat{B}^*$	
$\mathbf{s}_1$	$281367 \leqslant \mathbf{s}_1 \leqslant -$
$\mathbf{s}_2$	$4154920 \leqslant \mathbf{s}_2 \leqslant -$
Dist_Intersection	$19.6 \leqslant g_1 \leqslant 55.7$
Dist_TrafficLight	$16.5 \leqslant g_2 \leqslant -$
Dist_TrPark	$6219 \leqslant g_4 \leqslant -$
Dist_Shopping	$853 \leqslant g_5 \leqslant 5918$
Dist_School	$340 \leqslant g_6 \leqslant 680$
Dist_RrhaOwned	$- \leqslant g_7 \leqslant 1522$
Dist_PoliceSta	$- \leqslant g_9 \leqslant 2105$
Dist_PlWorship	$78.1 \leqslant g_{10} \leqslant 782$
Dist_Mall	$- \leqslant g_{12} \leqslant 4487$
Dist_Industrial	$- \leqslant g_{13} \leqslant 2926$
Dist_HistLand	$- \leqslant g_{14} \leqslant 1301$
Dist_CommunityCent	$335 \leqslant g_{15} \leqslant -$
Dist_Bridge	$- \leqslant g_{16} \leqslant 2656$

No socio-economic features were used by PRIM. All distances are in meters and the spatial coordinates are UTM.

For each simulation, the labels for the events are randomly assigned to τ_i, $(i = 1, 2)$, with $f(k = 1|\mathbf{s}) = 0.293$, and 500 iterations of PRIM are performed.

For the original observation, $\log t^{\text{obs}}$ was found to be -35.4175 in a region $\widehat{B}^*$ where there were 76 events from τ_1 and 36 events from τ_2. PRIM peeled 24 times on 15 different features to find this region. $\widehat{B}^*$ consisted of approximately 1.3% of the total area of A. In this case, 19 of the simulated observations were less than $\log t^{\text{obs}}$, so the estimated p-value of the observation is 0.019, and we can reject H_0 and conclude a change in the intensity has occurred in region $\widehat{B}^*$ between time periods τ_1 and τ_2. Fig. 6 shows $\widehat{B}^*$ as the shaded region in A. Table 3 shows the change region in feature space.

The change was caused by an elevation in the number of summertime crimes over what was expected given the non-summertime crimes in the identified region $\widehat{B^*}$. While under the null, $\theta(\mathbf{s}) = 0.4147$, the estimate of $\theta(\mathbf{s})$ for region $\widehat{B^*}$ is $\hat{\theta}(\mathbf{s} \mid \mathbf{s} \in B) = \hat{\theta}_{\mathrm{ML}} = 76/36 = 2.111$. This identifies the ability of this method to identify even small regions that constitute a change in the process. As seen from Fig. 6, when the change region is projected onto the geographic space, the change region is composed of several disjoint spatial regions and all of these subregions are in close spatial proximity to each other. While this could suggest the possibility that an additional group of criminals was active in this region during the summertime, this analysis alone cannot provide certain evidence. Although it would have been interesting to further investigate the possible causes of this change, this was unfortunately not possible. Certainly an evaluation of the crime reports of the B&E crimes that occurred inside the change region of $\widehat{B^*}$ or an examination of the arrests records from the B&E crimes in 1997 could help confirm this conjecture; unfortunately this was not possible.

However, this method does provide a list of crimes $\left(\left\{\mathbf{s}_i \in \widehat{B^*}\right\}\right)$, and regions $\widehat{B^*}$, in geographic space (Fig. 6) and feature space (Table 3), that can be further investigated by crime analysts to determine if a new criminal group was active during this period of time (or to test other possible hypotheses).

6. Discussion and conclusions

This method provides an approach for detecting local change regions in a stochastic point process on a plane where the locations of the points are driven by intelligent site selection. We have only presented examples where the changes occur between two time periods, but the same method would apply for detecting differences between two types of events or in case–control data (the marks signify the types of events rather than the time period).

We detect change regions in the feature space corresponding to the predictor variables that are thought to influence the criminal or terrorist process. The change regions can then be projected back onto geographic space as in Fig. 6. By searching for change in feature space in addition to the spatial coordinates, it is clear that this method can detect the change regions that may not be detected under a spatial analysis alone. An additional advantage is that no a priori knowledge on the size or geometry of the change region is required.

By only testing for one connected set of change, this test is very similar to the use of the traditional scan statistic for detecting spatial clustering (Kulldorff, 1997; Glaz and Zhang, 2004). In the traditional approach the search is over all possible connected sets of some geometry. Patil and Taillie (2004) show an alternative of finding the elevated regions where change has potentially occurred based on upper level sets. However, the events are required to be aggregated into a finite number of geographic cells for the analysis. The key difference here is that we have high dimensional, possibly mixed variable data which is not aggregated. As the number of events increase and especially as the number of variables increase, it becomes computationally infeasible to continue searching over every possible set. Therefore, in these occasions we propose the use of an approximation to the value of the scan statistic by means of the modified PRIM.

Furthermore, our test statistic (5) is very similar to those used in the spatial cluster detection methods. The GLRT used in spatial cluster detection compares the rate of event initiation inside the potential change region to the rate outside this change region. Since we are testing between two time periods (or types of events), we compare the rate of event initiation in a potential change region for one time period (or event type) to the rate in the same change region, but in the other time period (or event type).

By specifying the hypothesis test in terms of the ratio of mark densities, we are able to approach the inference problem by performing Monte Carlo simulations of the null process with random labeling, thus avoiding the complicated procedure and induced error of estimating the underlying intensities.

In this paper, we have concentrated on finding only one hyper-rectangular region (in feature space) of change. However, the PRIM approach could be extended to find other geometries or even multiple regions of change. After finding the first $\widehat{B^*}$, remove the events that fall inside this region. Then repeat PRIM resulting in another region, $\widehat{B^*}_{(2)}$. This can be continued until the space A is contained in the union of boxes or until some number or value of boxes is obtained. By combining overlapping regions, one connected but nonhyper-rectangular set can be identified for a possible change region. Or several disjoint regions can be tested as multiple change regions. Of course in this case, the hypothesis must be adjusted to incorporate the new assumptions.

When considering change detection, especially in the intelligent site section problem, it must be assumed that a change in the process could be in the form of new variables being considered. For example, in our example of crime or terrorism, the change might be in the form of a new decision maker that considers features that were not considered by anyone in the first time period. If any sort of feature selection is used, the features identified as relevant must come from the union of relevant features in each time period.

We are currently investigating extensions to this approach. Other potential application areas include anomaly detection, syndromic surveillance, and detecting changes in transportation patterns. Also, we have not considered the geographical features that vary temporally. However, such features could be very relevant in the analysis. For example, the actual locations and times of police or military patrols could have a profound effect on locations and number of crimes. We are examining the extension of this approach to capture the temporally varying geographic features.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G., 2000. Measurement and Analysis of Crime and Justice of Criminal Justice 2000. Spatial Analyses of Crime, first ed., vol. 4. National Institute of Justice, Washington, DC, pp. 213–258.

Besag, J., Newell, J., 1991. The detection of clusters in rare diseases. J. Roy. Statist. Soc.—Ser. A 154 (1), 143–155.

Brantingham, P.J., Brantingham, P., 1981. Environmental Criminology. Sage Publications, Beverly Hills, CA.

Bursik, R.J., 1988. Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26 (6), 519–551.

Byrne, J., Sampson, R., 1986. The Social Ecology of Crime. Springer, New York.

Census_CD+Maps, 1998. Version 2.0. GeoLytics Inc, East Brunswick, NJ.

Cohn, E.G., Felson, M., 1979. Social change and crime rate trends: a routine activity approach. Amer. Soc. Rev. 44, 588–608.

Cohn, E.G., Rotton, J., 2000. Weather, seasonal trends and property crimes in Minneapolis, 1987–1988. A moderator-variable time-series analysis of routine activities. J. Environmental Psychol. 20 (3), 257–272.

Cornish, L.E., Clarke, R., 1986. The Reasoning Criminal: Rational Choice Perspectives on Offending. Springer, New York.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Daley, D.J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes, vol. I. Springer, New York, NY.

Diggle, P., 1983. Statistical Analysis of Spatial Point Patterns. Academic Press, London, New York.

Fotheringham, A., Charlton, M., Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment Plann. 30 (11), 1905–1927.

Friedman, J.H., Fisher, N.I., 1999. Bump hunting in high-dimensional data. Statist. Comput. 9 (2), 123–143.

Glaz, J., Balakrishnan, N., 1999. Scan Statistics and Applications. Birkhauser, Boston.

Glaz, J., Zhang, Z., 2004. Multiple window discrete scan statistics. Journal of Applied Statistics 31 (8), 967–980.

Hipp, J.R., Bauer, D.J., Curran, P.J., Bollen, K.A., 2004. Crimes of opportunity or crimes of emotion? Testing two explanations of seasonal change in crime. Social Forces 82 (4), 1333–1372.

Karr, A.F., 1986. Point Processes and Their Statistical Inference. Marcel Dekker, New York.

Kornak, J., Irwin, M.E., Cressie, N., 2003. Technical Report, Department of Statistics Preprint No. 726, The Ohio State University.

Kulldorff, M., 1997. A spatial scan statistic. Communications in Statistics: Theory and Methods. 26 (6), 1481–1496.

Kulldorff, M., Tango, T., Park, P.J., 2003. Power comparisons for disease clustering tests. Comput. Statist. Data Anal. 42 (4), 665–684.

Law, J., Haining, R., 2004. A bayesian approach to modeling binary data: the case of high-intensity crime areas. Geographical Anal. 36 (3), 197–216.

Liu, H., Brown, D.E., 2003. Criminal incident prediction using a point-pattern based density model. Internat. J. Forecasting 19 (4), 603–622.

Loader, C., 1991. Large-deviation approximations to the distribution of scan statistics. Adv. in Appl. Probab. 23 (4), 751–771.

McFate, M., 2005. Iraq: the social context of IEDs. Military Rev. 85 (3), 37–40.

Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. CRC Press, Boca Raton, FL.

Priebe, C.E., Olson, T., Dennis, M., Healy, J., 1997. A spatial scan statistic for stochastic scan partitions. J. Amer. Statist. Assoc. 92 (440), 1476–1484.

Priebe, C.E., Naiman, D.Q., Cope, L.M., 2001. Importance sampling for spatial scan analysis: computing scan statistic p-values for marked point processes. Comput. Statist. Data Anal. 35 (4), 475–485.

Patil, G.P., Taillie, C., 2004. Upper level set scan statistics for detecting arbitrarily shaped hotspots. Environmental and Ecological Statistics. 11, 183–197.

Ratcliffe, J., 2005. Detecting spatial movement of intra-region crime patterns over time. J. Quantitative Criminol. 21 (1), 103–123.

Rogerson, P., 2001. Monitoring point patterns for the development of space–time clusters. J. Roy. Statist. Soc.—Ser. A 164 (1), 87–96.

Rogerson, P., Sun, Y., 2001. Spatial monitoring of geographic patterns: an application to crime analysis. Computers Environment Urban Systems 25 (6), 539–556.

Rossmo, D.K., 1999. Geographic Profiling. CRC Press, Boca Raton, FL.