Spatio-temporal classification in point patterns under the presence of clutter

Marianna Siino
Istituto Nazionale di Geofisica e Vulcanologia,
Centro Nazionale Terremoti, Rome, Italy
and
Francisco J. Rodríguez-Cortés
Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia
and
Jorge Mateu
Departament of Mathematics, University Jaume I, Castellón, Spain
and
Giada Adelfio*
Dipartimento di Scienze Economiche, Aziendali e Statistiche,
Università degli Studi di Palermo, Palermo, Italy

February 14, 2019

Abstract

We consider the problem of detection of features in the presence of clutter for spatio-temporal point patterns. In the spatial context, Byers and Raftery (1998) used K-th nearest-neighbour distances to classify points between clutter and features. They proposed a mixture of distributions whose parameters were estimated using an EM algorithm. This paper extends this methodology to the spatio-temporal context by considering the properties of the spatio-temporal K-th nearest-neighbour distances. For this purpose we make use of a couple of spatio-temporal distances which are based on the Euclidean and the maximum norms. We show close forms for the probability distributions of such K-th nearest-neighbour distances, and present an intensive simulation study together with an application to earthquakes.

Keywords: Clutter; Earthquakes; EM algorithm; Features; Mixtures; Nearest-neighbour distances; Spatio-temporal point patterns.

*Corresponding author: giada.adelfio@unipa.it (Giada Adelfio)

1 Introduction

One of the most important research fields of spatio-temporal data-mining is the identification of features (clusters) of events. In particular, features are defined as subgroups of events in restrained spatio-temporal volumes with a higher density than the events outside the spatio-temporal volumes (called background, noise or clutter events). The identification of such spatio-temporal features may yield insight for many applications. In practice, many geographical phenomena (e.g. earthquakes, disease cases, crime data, forest fires) are modelled as spatio-temporal events, and the detection of features is used to study the evolution of the phenomena, to reveal space or time anomalies and spatio-temporal hotspots (Ratcliffe, 2005; Demattei and Cucala, 2010; Izakian and Pedrycz, 2012; Eckley and Curtin, 2013; Costa and Kulldorff, 2014; Cheng and Wicks, 2014). However, the detection of features is a challenge problem for the complexity caused by the space-time coupling and the noise interference. Moreover, another element to consider depending on the used methodology is the subjectivity in the choice of crucial aspects such as the spatio-temporal window, and the corresponding critical values to classify an event as a feature or noise.

For point processes, the problem has been widely addressed from a spatial point of view. Allard and Fraley (1997) developed a method to find the maximum likelihood solution using Voronoi polygons. Dasgupta and Raftery (1998) used model-based clustering to extend the methodology proposed by Banfield and Raftery (1993). An alternative and flexible Bayesian methodology that is adaptable to these and other assumptions about the mine and clutter patterns was given by Cressie and Lawson (2000). Some of the methods described above are somewhat restricted. For instance, Allard and Fraley (1997) assumed that there is a single connected feature, whereas Dasgupta and Raftery (1998) assumed the features to have a specific shape. Byers and Raftery (1998) adopted a different approach in which they estimated and removed the clutter without making any assumptions about the shape or number of features. This enabled them to estimate features (and remove clutter) in quite general situations. In particular, their method of detection is based on the distance to the K-th nearest-neighbour of a point in a spatial process. The observed distances are modelled as a mixture distribution of distances coming from clutter and feature points, the parameters of which are estimated by an EM algorithm. Following the steps of Byers and Raftery (1998), Mateu et al. (2007) use distances between local product density functions, named LISA functions, for local indicator of spatial association, rather than K-th nearest-neighbour distances,

to remove clutter from images where one or several features are present and have to be detected.

The scan statistics are well established methodological tools to identify clusters and to assess their significance in case events or in comparisons of case-control data. They have been developed for both spatial (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995) and spatio-temporal data (Kulldorff et al., 1998, 2005; Kulldorff, 2001). Working with space-time data, the focus may be retrospective (Kulldorff et al., 1998; Assunçäo et al., 2007) or on prospective surveillance of space-time clusters (Kulldorff, 2001; Assunção and Correa, 2009). Moreover, in the spatio-temporal context Pei et al. (2010) have proposed a method for the identification of spatio-temporal clusters based on the windowed K-th nearest (WKN) distance. Demattei and Cucala (2010) proposed ordering-based methods used for detecting spatio-temporal clusters and highlighting the flexibility and their low computational demand. Liu et al. (2014) propose a novel spatio-temporal clustering method based on spatio-temporal shared nearest-neighbours for clusters with different sizes, shapes, and densities.

In this paper, we use spatio-temporal distances obtained as a weighted version with respect to the temporal component of the Euclidean and maximum distances (Demattei and Cucala, 2010) and they are described in Section 2. Moreover, it is proven that the distribution of the K-th nearest-neighbour based on the previous distances follow an inverse Gamma distribution under the homogeneous Poisson assumption. Based on the previous results, we extend in Section 3 the methodology of Byers and Raftery (1998) to the spatio-temporal context by considering the properties of the spatio-temporal K-th nearest-neighbour distances. In Section 4 the K-th nearest-neighbour distance is analysed through a mixture model formulation of the corresponding distance densities coming from the clutter and feature events. The corresponding parameters associated to the two density distributions in the mixture model formulation are estimated using an expectation-maximization (EM) algorithm. A simulation study with several scenarios is carried out in Section 5 to assess the performance of the proposed classification method. Our method is also compared with the results obtained with the spatial methodology of Byers and Raftery (1998) in terms of sensitivity, specificity and accuracy. Finally, we present in Section 6 a seismic application, identifying noise and feature events in two well-known seismic sequences occurred in California (near the Landers town, in 1992) and in Italy (L'Aquila city, in 2009).

2 Mathematical background

We consider a spatio-temporal point process with no multiple points as a random countable subset X of $\mathbb{R}^{d-1} \times \mathbb{R}$, where a point $(\mathbf{u}, s) \in X$ corresponds to an event at $\mathbf{u} \in \mathbb{R}^{d-1}$ occurring at time $s \in \mathbb{R}$. We observe n events $\{(\mathbf{u}_i, s_i)\}_{i=1}^{n}$ of distinct points of X within a bounded spatio-temporal region $W \times T \subset \mathbb{R}^{d-1} \times \mathbb{R}$, with volume $|W| > 0$, and with length $|T| > 0$ where $n \geq 0$ is not fixed in advance. In the sequel, $N(B)$ denotes the number of events of the process falling in a bounded region $B \subset W \times T$, for more details see (Diggle, 2013; González et al., 2016). We assume that the basic point process X is completely stationary, i.e. for all $\mathbf{u} \in \mathbb{R}^{d-1}$ and all real s it holds that $X \stackrel{\mathcal{D}}{=} X_{(\mathbf{u},s)}$, where $X_{(\mathbf{u},s)}$ is the translated process given by $X_{(\mathbf{u},s)} = \{[\mathbf{u}_1 + \mathbf{u}, s_1 + s], [\mathbf{u}_2 + \mathbf{u}, s_2 + s], \ldots\}$. In the context of summary characteristics we additionally assume complete isotropy, which means that all rotations of X around the origin of $\mathbb{R}^{d-1} \times \mathbb{R}$ have the same distribution as the original spatio-temporal point process X. A completely stationary spatio-temporal point process has a constant intensity λ, defined as $\lambda = \mathbb{E}[N(B)]$, i. e., λ is the mean number of points per unit volume and unit time (Illian et al., 2008; Ballani et al., 2018).

Let $(\mathbf{u}, s) = (u_1, u_2, \ldots, u_{d-1}, s)$, and $(\mathbf{v}, l) = (v_1, v_2, \ldots, v_{d-1}, l)$ be points of a spatio-temporal homogeneous Poisson process in $W \times T \subset \mathbb{R}^{d-1} \times \mathbb{R}$. Hereafter we denote by D^S the spatial $(d-1)$-dimensional distance defined by the p-norm as

$$D^S(\mathbf{u}, \mathbf{v}) = \left(\sum_{i=1}^{d-1} |u_i - v_i|^p \right)^{1/p} \quad \text{for } \mathbf{u}, \mathbf{v} \in \mathbb{R}^{d-1},$$

where $p \geq 1$ is a suitable real value to measure spatial distances induced by the p-norm (Bartle and Sherbert, 2000). In the same way, it is possible to introduce the temporal distance D^T given by

$$D^T(s, l) = |s - l| \text{ for } s, l \in \mathbb{R}.$$

Following Dematteiand Cucala (2010), we consider the spatio-temporal d-dimensional distance as a mixture of the defined distances D^S and D^T, given by the expression

$$D^{ST}((\mathbf{u}, s), (\mathbf{v}, l)) = \left(D^S(\mathbf{u}, \mathbf{v})^p + \rho^p D^T(s, l)^p \right)^{\frac{1}{p}},$$

where ρ is a scaling coefficient for the temporal and spatial scales to be commensurate. For the case $d = 3$ and $p = 2$, the Euclidean norm is obtained. Let $|W|$ be the area of the observation

region and $|T|$ the time interval length. Then, $D = 2\sqrt{|W|/\pi}$ is the diameter of a disc with area $|W|$ and therefore ρ is the ratio $D/|T|$. The spatio-temporal Euclidean distance for $d = 3$ and $p = 2$ is then given by

$$D^{ST_E}\left((\mathbf{u}, s), (\mathbf{v}, l)\right) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \rho^2 |s - l|^2}. \tag{1}$$

In addition, when $d = 3$ and $p = \infty$, D^{ST} is the well-known spatio-temporal maximum distance, which we denote as D^{ST_M} and is given by

$$D^{ST_M}\left((\mathbf{u}, s), (\mathbf{v}, l)\right) = \max\left(|u_1 - v_1|, |u_2 - v_2|, \rho|s - l|\right). \tag{2}$$

For more details about the spatio-temporal maximum distance and its usefulness in the spatio-temporal point process context, see Cronie and van Lieshout (2011).

3 K-nearest-neighbour distances and their distributions

Following Byers and Raftery (1998) and Mateu et al. (2007), we assume that the clutter is distributed as a homogeneous spatio-temporal Poisson point process in $W \times T \subset \mathbb{R}^{d-1} \times \mathbb{R}$, and the features are distributed as a Poisson process restricted to a certain spatio-temporal volume in $W \times T$ and overlaid on the clutter, defining a spatio-temporal Poisson process of piecewise constant rate. Considering a homogeneous spatio-temporal Poisson process in $W \times T$, the probability that there are no events within the hypersphere with radius ζ of an arbitrary point is given by $\mathbb{P}(D^{ST} \geq \zeta) = \mathbb{P}\left(N(B_\zeta) = 0\right) = \exp\{-\alpha_d \lambda \zeta^d\}$, where α_d is the coefficient in the volume of the hypersphere in the associated topology selected to the spatio-temporal distance, and given by $\text{Vol}(B_\zeta) = \alpha_d \zeta^d$. Therefore, the probability density function of D^{ST} (the point-to-nearest-event distance) is

$$f_{D^{ST}}(\zeta) = d\alpha_d \lambda \zeta^{d-1} \ \exp\{-\alpha_d \lambda \zeta^d\}, \tag{3}$$

for more details see Cressie (1993).

Furthermore, the probability that there are no events between distances ζ_{K-1} and ζ_K with $\zeta_{K-1} \leq \zeta_K$ of an arbitrary point in $W \times T$ is $\mathbb{P}\left(N(B_{\zeta_K} \setminus B_{\zeta_{K-1}}) = 0\right) = \exp\{-\alpha_d \lambda (\zeta_K^d - \zeta_{K-1}^d)\}$. The cumulative distribution and the probability density functions of $D_K^{ST} = \zeta_K$ (where D_K^{ST} is the distance from a random point in the process to its K-th nearest-neighbour), given $D_1^{ST} =$

$\zeta_1, D_2^{ST} = \zeta_2, \ldots, D_{K-1}^{ST} = \zeta_{K-1}$ for $0 \leq \zeta_1 \leq \zeta_2 \leq \cdots \leq \zeta_{K-1} \leq \zeta_K$, are

$$\begin{aligned} F_{D_K^{ST}}\left(\zeta_K | D_1^{ST} = \zeta_1, D_2^{ST} = \zeta_2, \ldots, D_{K-1}^{ST} = \zeta_{K-1}\right) &= 1 - \mathbb{P}\left(N(B_{\zeta_K} \setminus B_{\zeta_{K-1}}) = 0\right) \\ &= 1 - \exp\{-\alpha_d \lambda (\zeta_K^d - \zeta_{K-1}^d)\}, \end{aligned}$$

and

$$f_{D_K^{ST}}\left(\zeta_K | D_1^{ST} = \zeta_1, D_2^{ST} = \zeta_2, \ldots, D_{K-1}^{ST} = \zeta_{K-1}\right) = d\alpha_d \lambda \zeta_K^{d-1} \exp\{-\alpha_d \lambda (\zeta_K^d - \zeta_{K-1}^d)\}.$$

Therefore, the joint probability density function of $\zeta_1, \zeta_2, \ldots, \zeta_K$ is defined as

$$\begin{aligned} f_{D_1^{ST}, D_2^{ST}, \ldots, D_K^{ST}}(\zeta_1, \zeta_2, \ldots, \zeta_K) &= f_{D_1^{ST}}(\zeta_1) \times f_{D_2^{ST}}\left(\zeta_2 | D_1^{ST} = \zeta_1\right) \times \cdots \\ &\quad \times f_{D_K^{ST}}\left(\zeta_K | D_1^{ST} = \zeta_1, D_2^{ST} = \zeta_2, \ldots, D_{K-1}^{ST} = \zeta_{K-1}\right) \\ &= d\alpha_d \lambda \zeta_1^{d-1} \exp\{-\alpha_d \lambda \zeta_1^d\} \times d\alpha_d \lambda \zeta_2^{d-1} \exp\{-\alpha_d \lambda (\zeta_2^d - \zeta_1^d)\} \times \cdots \\ &\quad \times d\alpha_d \lambda \zeta_K^{d-1} \exp\{-\alpha_d \lambda (\zeta_K^d - \zeta_{K-1}^d)\}, \end{aligned}$$

and therefore we have that

$$f_{D_1^{ST}, D_2^{ST}, \ldots, D_K^{ST}}(\zeta_1, \zeta_2, \ldots, \zeta_K) = (d\alpha_d \lambda)^K \prod_{i=1}^{K} \zeta_i^{d-1} \exp\{-\alpha_d \lambda \zeta_K^d\}.$$

Consequently, the probability density function of ζ_K can be obtained by integrating the joint probability density function with respect to $\zeta_1, \zeta_2, \ldots, \zeta_{K-1}$ as

$$\begin{aligned} f_{D_K^{ST}}(\zeta_K) &= \int_0^{\zeta_{K-1}} \cdots \int_0^{\zeta_2} \int_0^{\zeta_1} f_{D_1^{ST}, D_2^{ST}, \ldots, D_K^{ST}}(\zeta_1, \zeta_2, \ldots, \zeta_K) \prod_{j=1}^{K-1} \mathrm{d}\zeta_j \\ &= (d\alpha_d \lambda)^K \exp\{-\alpha_d \lambda \zeta_K^d\} \int_0^{\zeta_{K-1}} \cdots \int_0^{\zeta_2} \int_0^{\zeta_1} \prod_{i=1}^{K} \zeta_i^{d-1} \prod_{i=j}^{K-1} \mathrm{d}\zeta_j. \end{aligned}$$

Additionally, using mathematical induction it is straightforward to proof that

$$\int_0^{\zeta_{K-1}} \cdots \int_0^{\zeta_2} \int_0^{\zeta_1} \prod_{i=1}^{K} \zeta_i^h \prod_{i=j}^{K-1} \mathrm{d}\zeta_j = \frac{\zeta_K^{(h+1)K-1}}{(h+1)^{K-1}(K-1)!}, \tag{4}$$

for some $h \in \mathbb{N}$.

Then, using (4) for $h = d - 1$ we get

$$\begin{aligned} f_{D_K^{ST}}(\zeta_K) &= (d\alpha_d\lambda)^K \exp\{-\alpha_d\lambda\zeta_K^d\} \int_0^{\zeta_{K-1}} \cdots \int_0^{\zeta_2} \int_0^{\zeta_1} \prod_{i=1}^{K} \zeta_i^{d-1} \prod_{i=j}^{K-1} \mathrm{d}\zeta_j \\ &= (d\alpha_d\lambda)^K \exp\{-\alpha_d\lambda\zeta_K^d\} \frac{\zeta_K^{dK-1}}{d^{K-1}(K-1)!} \\ &= \frac{d(\alpha_d\lambda)^K \exp\{-\alpha_d\lambda\zeta_K^d\}\zeta_K^{dK-1}}{(K-1)!} \\ &= \frac{d(\alpha_d\lambda)^K \exp\{-\alpha_d\lambda\zeta_K^d\}\zeta_K^{dK-1}}{\Gamma(K)}, \end{aligned}$$

where $\Gamma(\cdot)$ denotes the Gamma function and doing the change of variable $y = \zeta^d$ we obtain

$$f_{D_K^{ST}}\left(y^{1/d}\right) = \frac{d\alpha_d^K\lambda^K y^{K-1/d}\exp\{-\alpha_d\lambda y\}}{\Gamma(K)} \times \frac{y^{1/d-1}}{d} = \frac{(\alpha_d\lambda)^K y^{K-1}\exp\{-\alpha_d\lambda y\}}{\Gamma(K)},$$

which corresponds to a transformed Gamma random variable, $\left(D_K^{ST}\right)^d \sim \Gamma\left(K, \alpha_d\lambda\right)$. Then, D_K^{ST} follows an inverse Gamma distribution, namely

$$D_K^{ST} \sim \Gamma^{1/d}\left(K, \alpha_d\lambda\right). \tag{5}$$

In particular, the close form of the probability distribution function for the K-th nearest-neighbour Euclidean distance, under a homogeneous spatio-temporal Poisson process is given by

$$D_K^{ST_E} \sim \Gamma^{1/d}\left(K, \frac{\pi^{d/2}\lambda}{\Gamma\left(d/2+1\right)\rho}\right) \tag{6}$$

and that for the K-th nearest-neighbour maximum distance is

$$D_K^{ST_M} \sim \Gamma^{1/d}\left(K, \frac{2^d\lambda}{\rho}\right). \tag{7}$$

Given the values of D_K^{ST} from a homogeneous rate Poisson process, the maximum likelihood estimate (MLE) of the rate of the process is given by

$$\hat{\lambda} = \frac{K}{\alpha_d \sum_{i=1}^{N} \gamma_i^n}$$

where the γ_i are the observations of D_K^{ST} and N is the sample size.

4 Mixture Modelling

As in Byers and Raftery (1998) and Mateu et al. (2007), we assume to have two types of processes to be classified, and model the K-th nearest-neighbour distances through a mixture of the corresponding K-th nearest-neighbour distances coming from the clutter and feature, which are two superimposed spatio-temporal Possion processes. We then suppose that the distribution of the D_K^{ST} is roughly a mixture of distributions

$$D_K^{ST} \sim q\Gamma^{1/d}\left(K, \alpha_d\lambda_1\right) + (1-q)\,\Gamma^{1/d}\left(K, \alpha_d\lambda_2\right) \tag{8}$$

where λ_1 and λ_2 are the intensities of the two homogeneous spatio-temporal Poisson point processes (clutter and feature) and q is the constant which characterises the postulated distribution of the D_K^{ST}. The corresponding parameters associated to this mixture are estimated using an expectation-maximisation (EM) algorithm, where in the expectation step we use the close form provided by an inverse Gamma distribution. Each data point has an observation γ_i of D_K^{ST} and an unknown δ_i, where $\delta_i = 1$ if this point belongs to the feature, and $\delta_i = 0$ otherwise.

The expectation step ($\mathbb{E}$ step) of the algorithm consists of

$$\mathbb{E}\left(\hat{\delta}_i^{(z+1)}\right) = \frac{\hat{p}^{(z)} f_{D_K^{ST}}\left(\gamma_i; \hat{\lambda}_1^{(z)}\right)}{\hat{p}^{(z)} f_{D_K^{ST}}\left(\gamma_i; \hat{\lambda}_1^{(z)}\right) + \left(1-\hat{p}^{(z)}\right) f_{D_K^{ST}}\left(\gamma_i; \hat{\lambda}_2^{(z)}\right)}$$

and the maximisation step (M step) consists of

$$\hat{\lambda}_1^{(z+1)} = \frac{K\sum_{i=1}^N \hat{\delta}_i^{(z+1)}}{\alpha_d \sum_{i=1}^N \gamma_i^d \hat{\delta}_i^{(z+1)}}, \qquad \hat{\lambda}_2^{(z+1)} = \frac{K\sum_{i=1}^N \left(1-\hat{\delta}_i^{(z+1)}\right)}{\alpha_d \sum_{i=1}^N \gamma_i^d \left(1-\hat{\delta}_i^{(z+1)}\right)},$$

and

$$q^{(z+1)} = \frac{\sum_{i=1}^N \hat{\delta}_i^{(z+1)}}{N}.$$

Following Byers and Raftery (1998) the simplest test criterion is to classify according to the component of the mixture under which the observed D_K^{ST} has higher density. In our analysis, the entropy-type measure of separation $S = \sum_{i=1}^n \delta_i log(\delta_i)$, where δ_i is the probability that the $i-th$ point is in the first component of the mixture, is used to select the value of K (Byers and Raftery, 1998).

5 Simulation study

A simulation study is carried out to assess the performance of the proposed methodology in terms of detection of features in a spatio-temporal setting. The spatio-temporal classification procedure proposed in this paper is compared with the method based on the spatial K-th nearest-neighbour distance in Byers and Raftery (1998), named $M_{spatial}$, as if we ignore time. The spatio-temporal window is set as $W \times T = [0,1]^2 \times [0,50]$, where the time range is greater than the spatial one to have a different scale than the spatial window. In the spatio-temporal window, the clutters and features are simulated from two different processes, for a wide range of scenarios. In all of them, the clutter points are taken from a spatio-temporal homogeneous Poisson process with size $n_c = \{400, 1000\}$. Moreover, the shape and the number of features (clusters) change.

The number of feature points for each cluster is $n_{fc} = \{200, 400\}$, so the total number of feature points ($n_f = n_{fc} \times n_{cluster}$) changes according to the number of clusters ($n_{cluster}$).

Figure 1 shows various simulated point patterns for $n_c = 400$ clutter points and a number of feature points per cluster of $n_{fc} = 200$, for several shapes of the clusters. In particular, Figure 1a shows space-time feature points in a single ellipsoid, while we have three ellipsoids in Figure 1b. Figure 1c shows feature points defined only in space for three hotspots, while Figure 1d represents only temporal features based on two time layers.

The mixture model formulation depends on the type of distance D^{ST} and the corresponding K-th spatio-temporal nearest-neighbour distance. The two spatio-temporal distances defined in Section 2 are compared, namely the Euclidean (D^{ST_E}) and the Maxima (D^{ST_M}) distances. We report results for $K = \{5, 10\}$ and $\rho = \{0.02, 0.5, 1\}$. Parameter ρ rescales the temporal distance, such that the weight of the temporal term changes accordingly. When $\rho = 1$ and $D^{ST} = D^{ST_E}$, the methodology is equal to $M_{spatial}$ in three dimensions (Byers and Raftery, 1998). Instead, when $\rho = 0.02$, it corresponds to simulate the points in the unit cube. For each scenario, 100 point patterns are simulated as described above. Given a point pattern, each point is classified into feature or clutter on the basis of the mixture model formulation and the (EM) estimation algorithm detailed in Section 4.

The evaluation and comparison of the classification procedures with respect to the different settings are done in terms of the true positive rate (TPR, that is the proportion of feature points that are correctly identified as such), the false positive rate (FPR, that is the proportion of clutter

points that are wrongly identified as such) and the accuracy (the proportion of corrected classified points) (Stehman, 1997). For the features in space and time (e.g. Figure 1a and Figure 1b), the results are shown in Table 1 and Table 2, respectively; for the spatial clusters (e.g. Figure 1c), the results come in Table 3, and for the features in time, such as the example in Figure 1d, the results are shown in Table 4. When the clusters are in space-time (with both one and three features in the window, see Table 1 and Table 2), our method always outperforms $M_{spatial}$, in terms of TPR, FPR and accuracy. Decreasing the value of ρ, so making closer the unit measurements of the space and time dimensions, the TPR, the specificity (1-FPR) and accuracy increase. Comparing the two types of distances, there are no remarkable differences; indeed the slightly worse performance of the spatio-temporal K-nearest-neighbour based on the weighted maxima distance (D^{ST_M}) seems to be due to chance.

TABLE 1 HERE

TABLE 2 HERE

TABLE 3 HERE

TABLE 4 HERE

In the presence of hot-spots (Table 3), the performance of our proposed method, in terms of the calculated rates, improves as ρ reduces. However, since the feature points are homogeneous in time, the method $M_{spatial}$ outperforms the proposed procedure. Anyway, the two methods become comparable as ρ decreases. Conversely, when the features are just in time (Table 4), as expected, $M_{spatial}$ does not detect feature points and all the rates are around 50%. Moreover, in

this case, the method performance decreases as ρ decreases, since the features are homogeneous in space and the time component has a lower weight, see (1). In general, we observe that changing the value of K (from 5 to 10) the overall performance is comparable. Moreover, as expected, increasing the number of clutter points in $W \times T$ ($n_c = 1000$), the results are in general slightly worse than the cases with $n_c = 400$.

6 Application to seismic data

In this section, the proposed method is applied to seismic data. Since earthquakes can be viewed as a spatio-temporal pattern, the identification of clustered earthquakes provides key information on seismic dynamics. Well-studied statistical models are based on the idea that the seismicity can be considered as the sum of "background" earthquakes (caused by tectonic loading) and "triggered" earthquakes (caused by stress transfer), that tend to cluster in sequence close in time and space. For example, the epidemic type aftershock sequence (ETAS) model (Ogata, 1988) is based on this idea.

A cluster of earthquakes is formed by the main event of each sequence, its foreshocks and its aftershocks, that could occur before and after the mainshock, respectively. Background events are spontaneous earthquakes that do not trigger a sequence of aftershocks and because of this characteristic, space-time features of principal earthquakes (main and isolated events) are close to those of a Poisson process that is stationary in time, since the probability of occurrence of future events is constant in time irrespectively of the past activity, even if it is inhomogeneous in space. Therefore, the seismogenic features controlling the kind of seismic release of background and clustered seismicity are not similar (Adelfio et al., 2006), and to describe the seismicity of an area in space, time and magnitude domains, sometimes it is useful to study separately the features of *independent* events and *triggered* ones. For instance, to estimate parameters of phenomenological laws useful for the description of seismicity, a reasonable definition of "earthquake cluster" is required. Furthermore, the prediction of the occurrence of large earthquakes (related to the assessment of seismic risk in space and time) is complicated by the presence of clusters of aftershocks, that are superimposed to the background seismicity, according to some (unknown) mixing parameter, and shade its principal characteristics (Adelfio and Chiodi, 2015). For these purposes the preliminary subdivision of a seismic catalogue in

background seismicity (represented by isolated events, that do not trigger any further event) and clustered events, is many times required. At this regard, the proposed method based on the EM algorithm allows the identification of these two main components, such that the background seismicity is related to the long-term analysis, and the triggered one for sequence identification. Our goal is to detect spatio-temporal features (clusters) of earthquakes that occurred in central Italy and California. Choosing a spatial window around high seismic areas, we identified catalogue subsets where clearly a sequence of earthquakes occurred in a specific time period. The details of the spatial area, maximum depth, minimum magnitude and time period are reported in Table 5. In the centre of Italy, the study area is located near L'Aquila city, that was hit by a strong earthquake in April 4th, 2009. The earthquake was felt throughout central Italy, causing 308 deaths, making this one of the strongest earthquakes to hit Italy since the 1980 Irpinia earthquake. A total of 449 earthquakes with a magnitude greater than 3 were recorded in the selected area (see Figure 2). On the other hand, the subset of the earthquake catalogue in California refers to a study area located between -120–115 N and 32-36 E, near the Landers town, where in June 25^{th}, 1992 an event with magnitude 7.3 occurred, causing severe damage to the area directly surrounding the epicenter. A total number of 804 were observed with magnitude at least 3.5 from 1968 to 2012 (see Figure 4).

TABLE 5 HERE

Here, the clutter and feature points are identified using the spatio-temporal Euclidean weighted distance (D^{ST_E}) since in the simulation study it is shown to perform better than the spatio-temporal maximum weighted distance. For each spatio-temporal catalogue, the value of ρ is set as the ratio between the maximum spatial distance over the maximum temporal distance observed between the events, such that $\rho_{italy} = 0.092$ and $\rho_{california} = 0.042$. In Figures 3a and 5a, the entropy measure for each value of K for the two datasets is reported. For the Italy subset, we selected the value of $K = 12$ since after this value, the entropy measure can be considered constant. Instead, for the California catalogue, the distance to the 25-th nearest-neighbour is used in the EM classification method. The histograms of the selected K-th nearest-neighbour distances based on the Euclidean weighted distance show bi-modality (see

Figure 3b and Figure 5b) indicating that the mixture modelling approach is realistic. The results applying the EM classification procedure are reported in Figure 3c and Figure 5c, where we can see that the feature points are clearly identified. The total number of clutter and feature points were $n_c = 183$ and $n_f = 266$, respectively for the Italy data. Instead, for the California data, $n_c = 477=$ and $n_f = 327$ and three spatio-temporal features are identified.

In Table 6, we report the number of clutter (n_0) and feature points (n_1) obtained with the proposed classification method based on the K-th nearest-neighbour method with D^{ST_E}, for the estimated ρ in the two datasets. Additionally, we also report the number of points not properly identified using the spatial method of Byers and Raftery (1998), neglecting time. In particular, on the basis of the estimated classification, n_{1-0} is the number of feature points 'wrongly' classified as clutters, and n_{0-1} is the number of clutter points 'wrongly' classified as features. For both the datasets n_{0-1} is higher than n_{1-0}. However, for the Italian catalogue the misclassification is less evident than the Californian one. Indeed, the California earthquakes present a more complex structure than L'Aquila data, with multiple features and strong clusters in time and, therefore, neglecting the time dimension, clustered points are easily mixed up with the background events, confirming the utility of our proposed method in space-time for this kind of application.

TABLE 6 HERE

7 Conclusions

In this paper, we have presented a classification method for identifying regions with higher point densities (features) in a spatio-temporal context extending the procedure of Byers and Raftery (1998) that is based on the spatial K-th nearest-neighbour distance. We use the weighted spatio-temporal distance where the temporal term is scaled to account for the space-time coupling, proving that both the Euclidean and maximum distances with a weight in the temporal coordinate follow an Inverse Gamma distribution under the homogeneous Poisson assumption. Based on these results, a mixture model formulation for the K-th nearest-neighbour

distance of clutter and feature points is considered to perform a binary classification using an iterative EM algorithm. With a simulation study, where several cluster structures are designed (i.e. cluster in space and time, only hotspots in space and only features in time), the proposed methodology is compared with the spatial and spatio-temporal version of Byers and Raftery (1998), where in the spatio-temporal version the time is not scaled. The comparison is done in terms of the true positive rate, the false positive rate and the accuracy. In general, when weighting the temporal component in the distance measure, the results of the classification improve. In comparison to the spatial method, the performances change according to the type of cluster structure. In particular, when there are one or more features in space and time and only in time, our methodology outperforms the other one. As an example of application, in contexts where this methodology can provide interesting contributions, we try to identify background and triggered seismicity in well-known earthquake sequences occurred both in Italy and California, where the different local seismogenetic characteristics provide different clustering structure of data. The proposed approach performs well in both situations, also comparing with the other existing approaches.

However, this kind of methodology can not be used to identify different clusters, but to classify points as noise or feature points. Moreover, the value of K for the K-th nearest-neighbour must be specified by the user, but can be selected using the entropy measure. Another limit to point out is that our method is based on the assumption that the feature process follows a Poisson process superimposed on a noise process and that the intensity of each cluster is constant, that for big spatio-temporal windows could be a restrictive assumption. Thus, for future work, it would be interesting to improve the results refining the Poisson assumption by allowing for local regularities and interactions in the feature points.

Acknowledgement

This paper has been partially supported by the national grant of the Italian Ministry of Education University and Research (MIUR) for the PRIN-2015 program, 'Complex space-time modelling and functional analysis for probabilistic forecast of seismic events'.

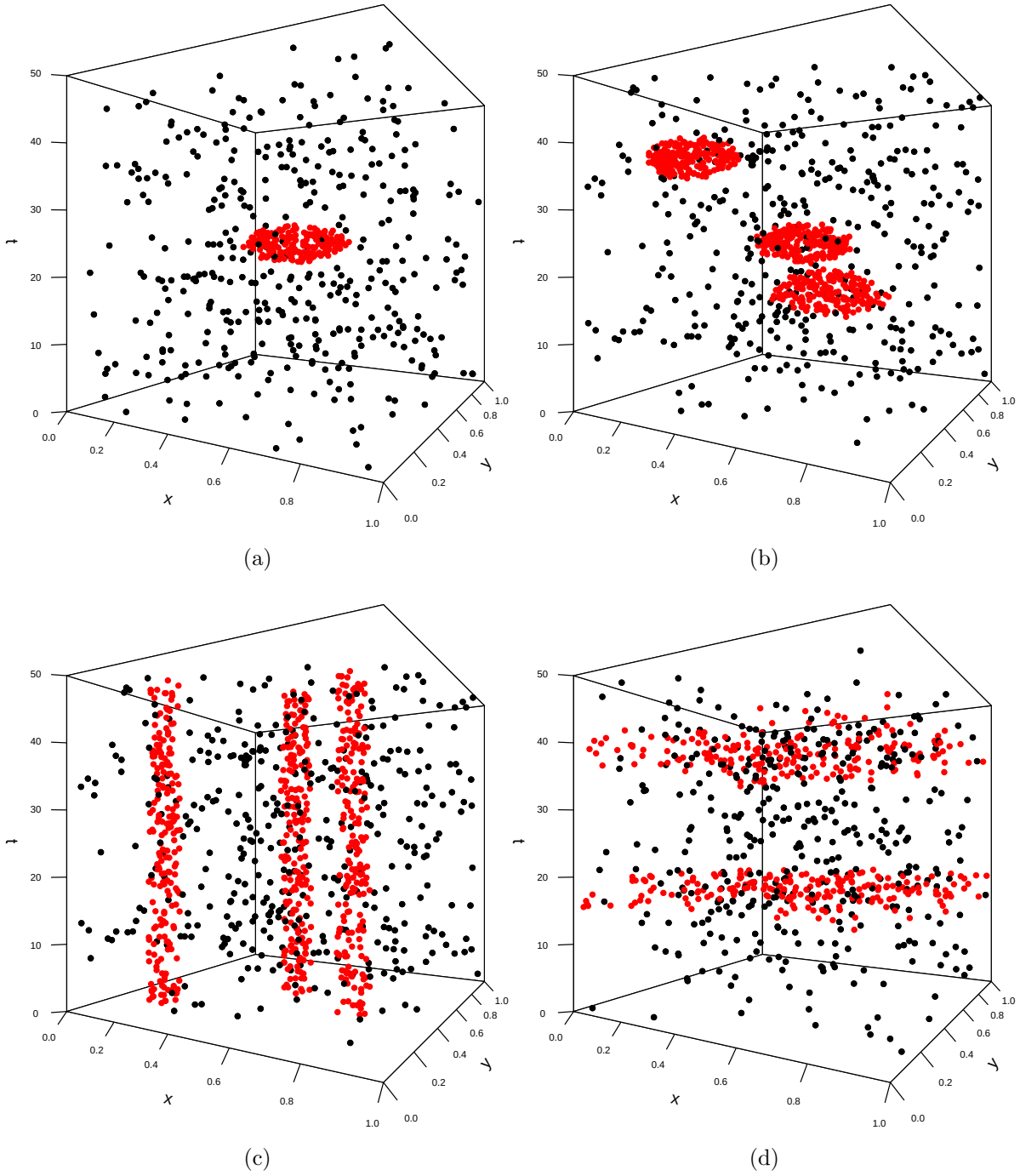
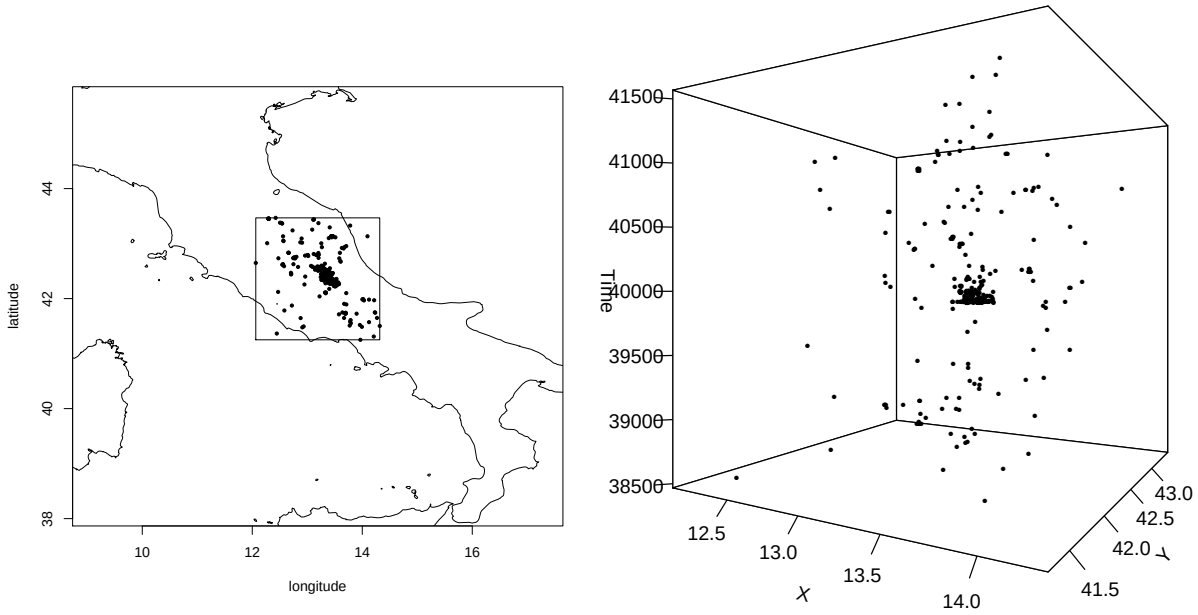
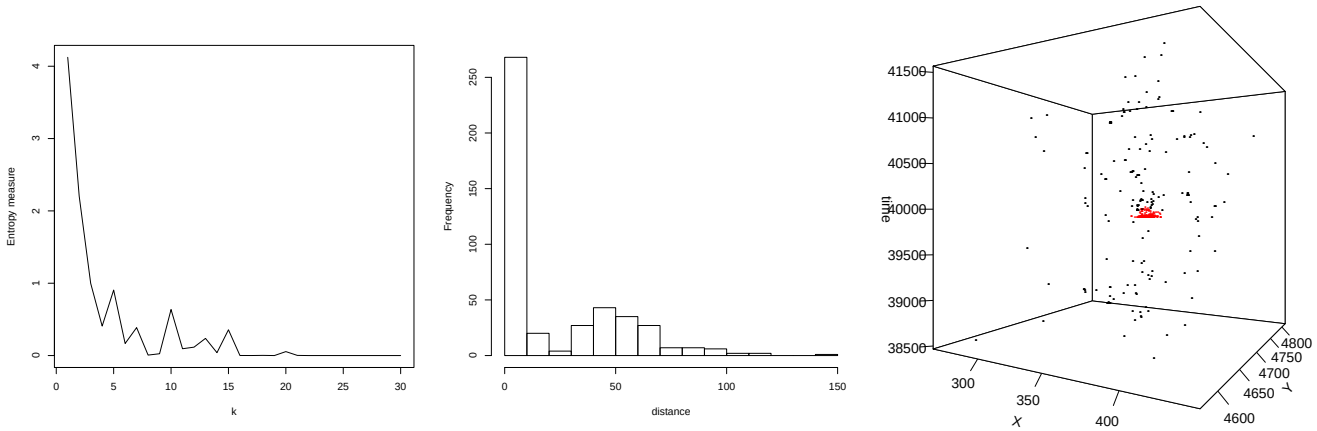


Figure 1: Simulated scenarios with $n_c = 400$ clutter points from a homogeneous Poisson point process in a spatio-temporal window $[0,1]^2 \times [0,50]$. In Figure 1a, $n_f = 200$ feature points are simulated in a centred ellipsoid with semi-axis $a = 0.2$, $b = 0.15$ and $c = 3.2$ with volume 0.41. In Figure 1b, there are three ellipsoids ($n_{clusters} = 3$) with semi-axis $a = 0.2$, $b = 0.15$ and $c = 3.5$ with volume 0.43, each with $n_{fc} = 200$ points, and $n_f = 600$. In Figure 1c, there are three cylindrical hotspots ($n_{clusters} = 3$) with radius 0.05, $n_{fc} = 200$ and $n_f = 600$. In Figure 1d, there are $n_{clusters} = 2$ in time each one with 200 points and $n_f = 400$.



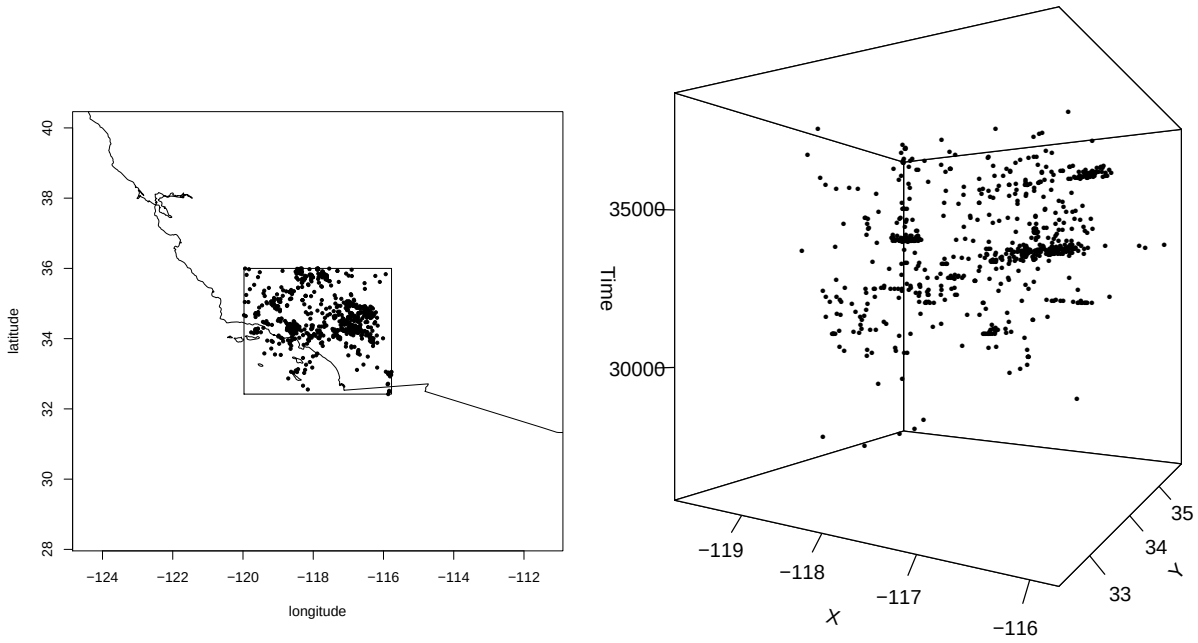
(a) (b)

Figure 2: 2D and 3D-plots of the earthquakes in the neighbourhood of L'Aquila, Italy.



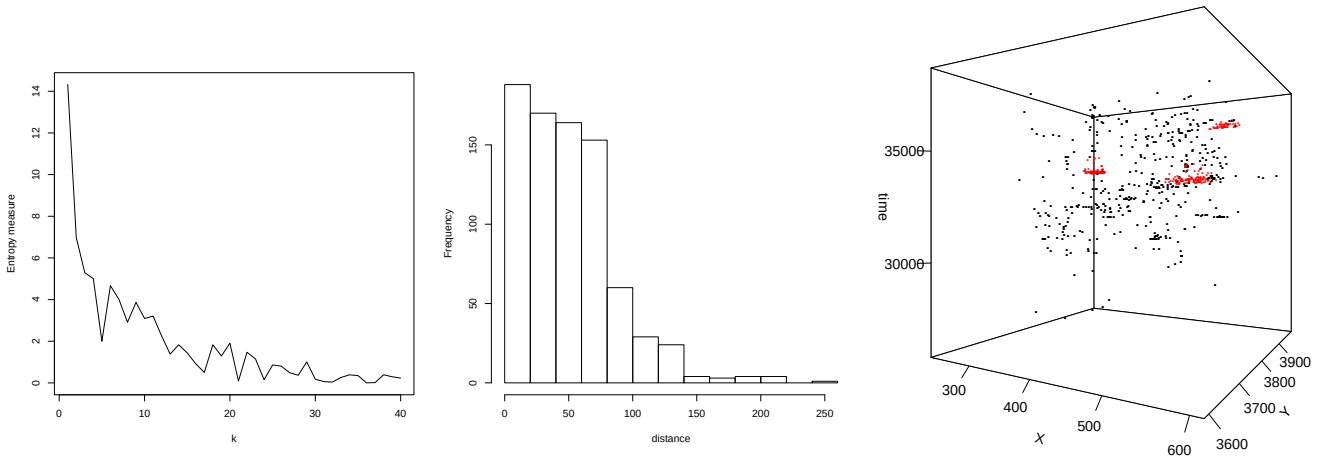
(a) (b) (c)

Figure 3: (3a) Entropy measure changing the K-th value of the K-th nearest-neighbour using D^{ST_E}. (3b) Histogram of the D^{ST_E} distance to the 12-th nearest-neighbour. (3c) Detected feature and clutter points with ρ_{italy} in D^{ST_E}.



(a) (b)

Figure 4: 2D and 3D-plots of the earthquake near Landers city, California.



(a) (b) (c)

Figure 5: (5a) Entropy measure changing the K-th value of the K-th nearest-neighbour using D^{ST_E}. (5b) Histogram of the D^{ST_E} distance to the 25-th nearest-neighbour. (5c) Detected feature and clutter points with $\rho_{california}$ in D^{ST_E}.

References

Adelfio, G. and M. Chiodi (2015). Flp estimation of semi-parametric models for spacetime point processes and diagnostic tools. *Spatial Statistics 14*, 119 – 132.

Adelfio, G., M. Chiodi, L. De Luca, D. Luzio, and V. M. (2006). Southern-tyrrhenian seismicity in space-time-magnitude domain. *Annals of Geophysics 49*(6), 1139–1151.

Allard, D. and C. Fraley (1997). Nonparametric maximum likelihood estimation of features in spatial point processes using voronoi tessellation. *Journal of the American Statistical Association 92*(440), 1485–1493.

Assunção, R. and T. Correa (2009). Surveillance to detect emerging space–time clusters. *Computational Statistics & Data Analysis 53*(8), 2817–2830.

Assunçäo, R., A. Tavares, T. Correa, and M. Kulldorff (2007). Space-time cluster identification in point processes. *Canadian Journal of Statistics 35*(1), 9–25.

Ballani, F., F. J. Rodríguez-Cortés, J. Mateu, and D. Stoyan (2018). Mark-based second-order characteristics in the statistics for spatio-temporal point processes. *Submitted*.

Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 803–821.

Bartle, R. G. and D. R. Sherbert (2000). *Introduction to Real Analysis*, Volume 2. Wiley New York.

Byers, S. and A. E. Raftery (1998). Nearest-neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association 93*(442), 577–584.

Cheng, T. and T. Wicks (2014). Event detection using twitter: a spatio-temporal approach. *PloS one 9*(6), e97807.

Costa, M. A. and M. Kulldorff (2014). Maximum linkage space-time permutation scan statistics for disease outbreak detection. *International Journal of Health Geographics 13*(1), 1–20.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Cressie, N. and A. B. Lawson (2000). Hierarchical probability models and bayesian analysis of mine locations. *Advances in Applied Probability 32*(2), 315–330.

Cronie, O. and M. N. M. van Lieshout (2011). A J-function for inhomogeneous point processes. *Statistica Neerlandica 65*(2), 183–201.

Dasgupta, A. and A. E. Raftery (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association 93*(441), 294–302.

Demattei, C. and L. Cucala (2010). Multiple spatio-temporal cluster detection for case event data: An ordering-based approach. *Communications in Statistics - Theory and Methods 40*(2), 358–372.

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press.

Eckley, D. C. and K. M. Curtin (2013). Evaluating the spatiotemporal clustering of traffic incidents. *Computers, Environment and Urban Systems 37*, 70–81.

González, J. A., F. J. Rodríguez-Cortés, O. Cronie, and J. Mateu (2016). Spatio-temporal point process statistics: A review. *Spatial Statistics 18*, 505 – 544.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, Volume 70. John Wiley & Sons.

Izakian, H. and W. Pedrycz (2012). A new pso-optimized geometry of spatial and spatio-temporal scan statistics for disease outbreak detection. *Swarm and Evolutionary Computation 4*, 1–11.

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

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 164*(1), 61–72.

Kulldorff, M., W. F. Athas, E. J. Feurer, B. A. Miller, and C. R. Key (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in los alamos, new mexico. *American Journal of Public Health 88*(9), 1377–1380.

Kulldorff, M., R. Heffernan, J. Hartman, R. Assunçao, and F. Mostashari (2005). A space–time permutation scan statistic for disease outbreak detection. *Plos med 2*(3), e59.

Kulldorff, M. and N. Nagarwalla (1995). Spatial disease clusters: detection and inference. *Statistics in Medicine 14*(8), 799–810.

Liu, Q., M. Deng, J. Bi, and W. Yang (2014). A novel method for discovering spatio-temporal clusters of different sizes, shapes, and densities in the presence of noise. *International Journal of Digital Earth 7*(2), 138–157.

Mateu, J., G. Lorenzo, and E. Porcu (2007). Detecting features in spatial point processes with clutter via local indicators of spatial association. *Journal of Computational and Graphical Statistics 16*(4), 968–990.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association 83*(401), 9–27.

Pei, T., C. Zhou, A.-X. Zhu, B. Li, and C. Qin (2010). Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise. *International Journal of Geographical Information Science 24*(6), 925–948.

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

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment 62*(1), 77–89.

Table 1: Results in terms of true positive rate (first row), false positive rate (second row) and accuracy (third row) in percentage over 100 simulated point patterns when the spatio-temporal cluster is an ellipsoid. An example is the point pattern in Figure 1a with $n_c = \{400\}$ and $n_{fc} = \{200\}$. The total number of clutter points is $n_c = \{400, 1000\}$. $n_{fc} = \{200, 400\}$ indicates the number of points for each feature, K is the K-th nearest-neighbour distance ($K = \{5, 10\}$) and $\rho = \{1, 0.5, 0.02\}$ is the weight for the temporal term in the spatio-temporal distance. D^{ST_E} and D^{ST_M} refer to different types of space-time distances. $M_{spatial}$ indicates the results obtained with the spatial method of Byers and Raftery (1998), neglecting time.

			$K = 5$			$K = 10$		
n_c	n_{fc}	ρ	D^{ST_E}	D^{ST_M}	$M_{spatial}$	D^{ST_E}	D^{ST_M}	$M_{spatial}$
400	200	1	97.96	97.39	96.82	97.14	96.38	97.81
			3.07	3.42	11.95	4.35	5.14	9.56
			97.27	96.85	90.98	96.15	95.36	92.89
		0.5	99.14	98.82	-	98.53	97.88	-
			2.08	2.29	-	2.78	3.04	-
			98.33	98.08	-	97.66	97.27	-
		0.02	99.86	99.82	-	99.96	99.92	-
			1.38	1.44	-	1.69	1.74	-
			99.03	98.98	-	98.86	98.81	-
	400	1	99.27	99.00	97.85	98.57	98.11	98.86
			2.12	2.32	9.35	2.93	3.30	9.10
			98.58	98.34	94.25	97.82	97.41	94.88
		0.5	99.70	99.60	-	99.43	99.13	-
			1.58	1.67	-	2.04	2.21	-
			99.06	98.97	-	98.70	98.46	-
		0.02	99.96	99.95	-	100.00	100.00	-
			1.22	1.25	-	1.44	1.50	-
			99.37	99.35	-	99.28	99.25	-
1000	200	1	97.87	97.42	95.75	97.00	96.37	96.04
			3.67	4.10	29.14	3.65	4.67	16.86
			96.59	96.16	75.01	96.46	95.50	85.29
		0.5	98.82	98.60	-	98.35	97.92	-
			2.62	2.74	-	2.46	2.71	-
			97.62	97.49	-	97.68	97.40	-
		0.02	99.70	99.60	-	99.84	99.78	-
			1.66	1.81	-	1.57	1.64	-
			98.57	98.42	-	98.66	98.60	-
	400	1	99.00	98.67	96.92	98.38	97.82	97.10
			2.20	2.30	14.32	2.66	2.92	9.53
			98.14	97.98	88.89	97.64	97.29	92.37
		0.5	99.50	99.36	-	99.22	98.99	-
			1.68	1.75	-	1.98	2.09	-
			98.65	98.57	-	98.37	98.22	-
		0.02	99.90	99.85	-	99.98	99.95	-
			1.22	1.25	-	1.40	1.44	-
			99.10	99.06	-	98.99	98.96	-

Table 2: Results in terms of true positive rate (first row), false positive rate (second row) and accuracy (third row) in percentage over 100 simulated point patterns when the spatio-temporal feature points are three ellipsoids. An example is the point pattern in Figure 1b with $n_c = \{400\}$ and $n_{fc} = \{200\}$. The total number of clutter points is $n_c = \{400, 1000\}$. $n_{fc} = \{200, 400\}$ indicates the number of points for each feature, K is the K-th nearest-neighbour distance ($K = \{5, 10\}$) and $\rho = \{1, 0.5, 0.02\}$ is the weight for the temporal term in the spatio-temporal distance. D^{ST_E} and D^{ST_M} refer to different types of space-time distances. $M_{spatial}$ indicates the results obtained with the spatial method of Byers and Raftery (1998), neglecting time.

			$K = 5$			$K = 10$		
n_c	n_{fc}	ρ	D^{ST_E}	D^{ST_M}	$M_{spatial}$	D^{ST_E}	D^{ST_M}	$M_{spatial}$
400	200	1	97.72	97.06	96.24	96.16	95.10	97.80
			8.82	9.66	26.89	11.89	13.74	26.42
			95.11	94.37	86.99	92.94	91.57	88.12
		0.5	98.97	98.60	-	98.03	97.46	-
			6.64	7.12	-	8.33	9.03	-
			96.73	96.31	-	95.49	94.87	-
		0.02	99.83	99.80	-	99.96	99.91	-
			4.80	4.97	-	5.72	5.89	-
			97.98	97.89	-	97.69	97.59	-
	400	1	99.08	98.77	97.75	98.31	97.74	98.78
			7.50	8.00	26.00	9.72	10.82	26.43
			97.43	97.08	91.81	96.30	95.60	92.48
		0.5	99.62	99.46	-	99.26	98.95	-
			5.81	6.13	-	7.26	7.80	-
			98.26	98.06	-	97.63	97.27	-
		0.02	99.96	99.94	-	100.00	99.99	-
			4.46	4.60	-	5.28	5.36	-
			98.86	98.81	-	98.68	98.66	-
1000	200	1	97.51	97.15	92.87	96.50	95.54	95.26
			9.20	10.10	33.07	10.43	11.72	27.44
			93.32	92.62	76.66	92.17	91.00	81.07
		0.5	98.59	98.32	-	98.20	97.73	-
			6.76	7.20	-	7.59	8.25	-
			95.25	94.87	-	94.58	93.99	-
		0.02	99.54	99.50	-	99.88	99.83	-
			4.49	4.63	-	5.05	5.22	-
			97.02	96.92	-	96.80	96.67	-
	400	1	98.79	98.50	95.64	98.09	97.51	97.04
			6.65	7.07	26.93	8.12	8.82	25.01
			96.32	95.97	85.38	95.27	94.63	87.02
		0.5	99.43	99.29	-	99.05	98.79	-
			5.37	5.63	-	6.39	6.76	-
			97.25	97.05	-	96.58	96.27	-
		0.02	99.83	99.83	-	99.97	99.96	-
			4.04	4.17	-	4.61	4.73	-
			98.07	98.01	-	97.89	97.83	-

Table 3: Results in terms of true positive rate (first row), false positive rate (second row) and accuracy (third row) in percentage over 100 simulated point patterns when the feature points are in space only and there are three hot spots. An example is the point pattern in Figure 1c with $n_c = \{400\}$ and $n_{fc} = \{200\}$. The total number of clutter points is $n_c = \{400, 1000\}$. $n_{fc} = \{200, 400\}$ indicates the number of points for each feature, K is the K-th nearest-neighbour distance ($K = \{5, 10\}$) and $\rho = \{1, 0.5, 0.02\}$ is the weight for the temporal term in the spatio-temporal distance. D^{ST_E} and D^{ST_M} refer to different types of space-time distances. $M_{spatial}$ indicates the results obtained with the spatial method of Byers and Raftery (1998), neglecting time.

			$K = 5$			$K = 10$		
n_c	n_{fc}	ρ	D^{ST_E}	D^{ST_M}	$M_{spatial}$	D^{ST_M}	D^{ST_M}	$M_{spatial}$
400	200	1	62.66	67.19	99.93	60.69	59.69	99.99
			34.11	37.13	3.21	38.21	36.86	3.58
			63.96	65.46	98.68	61.13	61.07	98.56
		0.5	77.17	67.41	-	73.00	81.72	-
			33.14	33.71	-	37.88	42.28	-
			73.05	66.96	-	68.65	72.12	-
		0.02	99.69	99.66	-	99.77	99.75	-
			6.04	6.40	-	7.42	7.94	-
			97.40	97.24	-	96.89	96.67	-
	400	1	73.05	55.13	99.98	64.65	81.02	100.00
			29.90	26.75	3.07	31.45	39.30	3.38
			72.31	59.66	99.22	65.62	75.94	99.15
		0.5	84.94	76.41	-	88.54	79.78	-
			23.03	26.07	-	37.94	39.90	-
			82.94	75.79	-	81.92	74.86	-
		0.02	99.90	99.90	-	99.97	99.96	-
			5.36	5.64	-	6.38	6.77	-
			98.59	98.51	-	98.38	98.28	-
1000	200	1	72.37	69.46	99.60	70.38	73.87	99.95
			41.28	42.20	3.02	43.75	42.97	3.22
			63.84	62.17	97.96	61.55	63.35	97.97
		0.5	80.67	75.40	-	84.34	81.00	-
			36.55	38.83	-	46.19	45.84	-
			69.91	66.51	-	65.26	64.23	-
		0.02	99.11	99.06	-	99.62	99.56	-
			5.57	5.91	-	6.34	6.87	-
			96.18	95.95	-	95.90	95.54	-
	400	1	78.97	69.97	99.91	75.86	80.32	100.00
			33.96	34.90	2.87	39.05	41.83	3.08
			73.10	67.75	98.65	69.08	70.25	98.60
		0.5	87.80	83.04	-	93.09	89.40	-
			23.95	28.59	-	42.31	47.53	-
			82.46	77.76	-	77.00	72.61	-
		0.02	99.70	99.69	-	99.89	99.88	-
			4.73	4.95	-	5.57	5.89	-
			97.69	97.58	-	97.41	97.25	-

Table 4: Results in terms of true positive rate (first row), false positive rate (second row) and accuracy (third row) in percentage over 100 simulated point patterns when the feature points are only in time and there are two clusters. An example is the point pattern in Figure 1d with $n_c = \{400\}$ and $n_{fc} = \{200\}$. The total number of clutter points is $n_c = \{400, 1000\}$. $n_{fc} = \{200, 400\}$ indicates the number of points for each feature, K is the K-th nearest-neighbour distance ($K = \{5, 10\}$) and $\rho = \{1, 0.5, 0.02\}$ is the weight for the temporal term in the spatio-temporal distance. D^{ST_E} and D^{ST_M} refer to different types of space-time distances. $M_{spatial}$ indicates the results obtained with the spatial method of Byers and Raftery (1998), neglecting time.

			$K = 5$			$K = 10$		
n_c	n_{fc}	ρ	D^{ST_E}	D^{ST_M}	$M_{spatial}$	D^{ST_E}	D^{ST_M}	$M_{spatial}$
400	200	1	95.00	95.16	54.15	97.59	98.15	61.50
			30.07	31.12	54.70	26.70	28.09	61.45
			82.46	82.02	49.72	85.45	85.03	50.03
		0.5	94.36	94.26	-	95.92	95.74	-
			28.57	29.02	-	25.12	25.75	-
			82.89	82.62	-	85.40	84.99	-
		0.02	89.64	89.28	-	89.83	88.12	-
			38.73	41.41	-	43.51	47.02	-
			75.46	73.94	-	73.16	70.55	-
	400	1	97.84	97.75	54.13	99.03	99.13	60.44
			24.56	24.64	54.05	22.16	22.69	60.04
			90.38	90.29	51.40	91.96	91.86	53.62
		0.5	97.52	97.34	-	98.45	98.36	-
			23.62	23.71	-	22.09	22.44	-
			90.47	90.32	-	91.60	91.43	-
		0.02	95.62	95.62	-	95.43	94.88	-
			34.80	37.34	-	38.55	42.30	-
			85.48	84.63	-	84.11	82.49	-
1000	200	1	89.07	88.78	53.12	91.24	90.84	58.24
			40.91	40.88	53.49	35.37	35.73	58.18
			67.66	67.60	48.40	72.23	71.86	46.51
		0.5	89.28	89.50	-	91.45	90.97	-
			39.98	40.86	-	35.39	35.91	-
			68.38	67.82	-	72.28	71.77	-
		0.02	84.86	85.02	-	88.51	87.96	-
			44.25	46.16	-	48.60	51.23	-
			64.07	62.75	-	62.00	59.96	-
	400	1	94.27	94.19	53.24	95.59	95.50	56.69
			30.81	31.20	53.23	25.39	25.96	56.43
			80.34	80.08	49.65	83.94	83.58	49.40
		0.5	94.26	94.14	-	95.37	95.09	-
			29.41	30.01	-	24.71	25.07	-
			81.11	80.72	-	84.21	83.89	-
		0.02	91.67	91.70	-	93.39	93.24	-
			36.46	38.52	-	38.13	40.97	-
			76.04	74.91	-	75.88	74.23	-

	Δx	Δy	h^*	Δt	m_0	n	max Km	max days
(a) L'Aquila (Italy)	12 - 14.5	41 - 43.5	30	2005 - 2013	3	449	286	3100
(b) Landers (California)	-120 - -115	32 - 36	95	1968 - 2012	3.5	804	658	7870

Table 5: Main information of the Italy (a) and California (b) catalogue subsets in terms of longitude (Δx), latitude (Δy), and time ranges (Δt). The minimum magnitude value is m_0, the maximum value for the depth is h^* and the corresponding number of selected events is n. It is also shown the maximum distance in kilometres and days between the events, and the value of the weight ρ used in the Euclidean distance.

		D^{ST_E}		$M_{spatial}$	
	$\hat{\rho}$	n_0	n_1	n_{0-1}	n_{1-0}
Italy	0.092	186 (41.43%)	263 (58.57%)	43 (9.57%)	4 (0.89%)
California	0.042	477 (59.33%)	327 (40.67%)	176 (21.89%)	18 (2.23%)

Table 6: Number of clutter (n_0) and feature points (n_1) obtained with the classification method based on the K-th nearest-neighbour method with the weighted D^{ST_E}. In the third and fourth columns, we show the classification obtained with the spatial method of Byers and Raftery (1998), neglecting time.