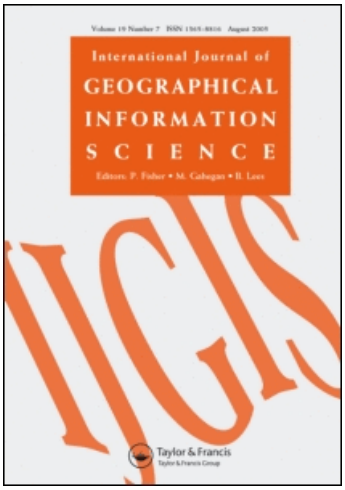

International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713599799

Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise

Tao Pei [a]; Chenghu Zhou [a]; A-Xing Zhu [a]; Baolin Li [a];Chengzhi Qin [a]
[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing, China

Online publication date: 16 April 2010

To cite this Article Pei, Tao , Zhou, Chenghu , Zhu, A-Xing , Li, Baolin andQin, Chengzhi(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
To link to this Article: DOI: 10.1080/13658810903246155
URL: http://dx.doi.org/10.1080/13658810903246155

International Journal of Geographical Information Science
Vol. 24, No. 6, June 2010, 925–948

Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise

Tao Pei, Chenghu Zhou*, A-Xing Zhu, Baolin Li and Chengzhi Qin

State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing, China

(*Received 11 May 2009; final version received 10 August 2009*)

In a spatio-temporal data set, identifying spatio-temporal clusters is difficult because of the coupling of time and space and the interference of noise. Previous methods employ either the window scanning technique or the spatio-temporal distance technique to identify spatio-temporal clusters. Although easily implemented, they suffer from the subjectivity in the choice of parameters for classification. In this article, we use the windowed *k*th nearest (WKN) distance (the geographic distance between an event and its *k*th geographical nearest neighbour among those events from which to the event the temporal distances are no larger than the half of a specified time window width [TWW]) to differentiate clusters from noise in spatio-temporal data. The windowed nearest neighbour (WNN) method is composed of four steps. The first is to construct a sequence of TWW factors, with which the WKN distances of events can be computed at different temporal scales. Second, the appropriate values of TWW (i.e. the appropriate temporal scales, at which the number of false positives may reach the lowest value when classifying the events) are indicated by the local maximum values of densities of identified clustered events, which are calculated over varying TWW by using the expectation-maximization algorithm. Third, the thresholds of the WKN distance for classification are then derived with the determined TWW. In the fourth step, clustered events identified at the determined TWW are connected into clusters according to their density connectivity in geographic–temporal space. Results of simulated data and a seismic case study showed that the WNN method is efficient in identifying spatio-temporal clusters. The novelty of WNN is that it can not only identify spatio-temporal clusters with arbitrary shapes and different spatio-temporal densities but also significantly reduce the subjectivity in the classification process.

Keywords: nearest neighbour; DBSCAN; cluster; spatio-temporal; windowed; expectation-maximization

1. Introduction

In a temporal–spatial event set, clusters (features) are referred to as subgroups of events in restrained spatio-temporal volumes whose densities are denser than events outside the volumes (background events or noise). The identification of spatio-temporal clusters may help with revealing the evolving patterns of spatial anomalies or locating the varying areas of temporal anomalies or detecting the time–space-coupled hot spots. Hence, the spatio-temporal clustering method, known as one of the important branches of spatial data mining and knowledge discovery, has been extensively examined and widely used in epidemiology, crime behaviour

*Corresponding author. Email: zhouch@lreis.ac.cn

ISSN 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/13658810903246155
http://www.informaworld.com

prediction, seismicity research, and so on (Sabel *et al.* 2000, Johnson and Bowers 2004, Bastin *et al.* 2007, Lian *et al.* 2007, Grubesic and Mack 2008). Because of their broad application area, it is very important to develop highly efficient spatio-temporal clustering methods.

Although very important, two obstacles exist in the identification of spatio-temporal clusters on account of the noise interference and the complexity caused by time–space coupling. The first is the determination of the membership of an event in spatio-temporal data, which means classifying events into feature and noise. The crucial issue in the first is the choice of the unit for the estimation of spatio-temporal density around an event (either cell statistics or distance technique was used in recent approaches, which will be reviewed later), which is used as a standard to classify events. The second is how to group spatial–temporal 'close' events into clusters, which depends on the mechanism for connecting feature events. Although many methods were developed to identify clusters in spatio-temporal event sets, subjectivities in the choice of thresholds (either with cell statistics or distance techniques) still remain, which may cause different classification results in terms of cluster number, cluster shape and the number of false positives (i.e. the feature events or noise misclassified as noise or feature events, respectively).

So far, clustering methods that were used to discover dense regions of events on the basis of the notion of density [namely, density-based method (Han *et al.* 2001)] are broadly classified into two groups. The first is the cell-based method and the second is the distance-based method. The cell-based methods first aggregate events into user-defined grid and then identify significantly clustering regions, which are made up of connected and dense cells. The cell-based methods for geographic clustering have been extensively investigated, for example, STING, CLIQUE and MAFIA (Wang *et al.* 1997, Agrawal *et al.* 1998, Nagesh *et al.* 2009). However, these methods can only deal with spatial objects. Another important approach is that of the scanning methods, such as the well-known SaTScan method (Kulldorff 1997), which is proposed for identifying the spatial clustering patterns that have changed with time; it has been widely used in disease surveillance (Kulldorff and Nagarwalla 1995, Kulldorff *et al.* 2005). In the SaTScan method, the scanning window, defined as a circle (with a spatial radius) or cylinder (with a circular geographic base and the height corresponding to time), is moved in space and time to detect significantly clustering regions (Kulldorff and Nagarwalla 1995, Kulldorff *et al.* 2005, Gaudart *et al.* 2008). Differing from the method we will discuss in this article, the scanning methods intend to identify truly significant clusters contained in pre-defined regions by excluding 'false clusters' (those are likely to have occurred by chance). As a result, the methods require simulated statistics, such as the Kulldorff scan statistics (Dwass 1957, Kulldorff 1997, Mostashari *et al.* 2003), or prior knowledge of the distribution of the events over the region of interest with respect to varying baselines (such as the distribution of Bernoulli, Poisson or exponential) (Gangnon 2006, Yan and Clayton 2006). Nevertheless, they are incapable of providing precise information of clusters in terms of position and shape. Moreover, the subjectivity in defining the window and target regions may cause significant different clustering results. Although the kernel estimation was proposed later as an alternative for the scanning method (Kelsall and Diggle 1995), it is not in itself a technique for detecting clustering. Instead, it can only generate distribution maps of density for further analysis.

The distance-based methods utilize the kth nearest distance, defined as the distance between an event and its kth nearest neighbour, to estimate local density instead of using cell-based statistics. Based on the relationship between the local density and the kth nearest distance, a variety of clustering models have been constructed for identifying spatial clustering patterns, such as DBSCAN, DENCLUE OPTICS, CHAMELEON and DECODE (Ester *et al.* 1996, Hinneburg and Keim 1998, Ankerst *et al.* 1999, Karypis *et al.* 1999, Pei *et al.* 2006, 2009). The

distance-based methods are less subjective in estimating local density because they do not need to specify the size and shape of the cell. However, the methods mentioned above are limited to spatial data. To extend the distance-based methods to spatio-temporal data, methods based on spatio-temporal nearest distance were developed, such as Knox test (Knox 1964, Williams 1984, Kulldorff and Hjalmars 1999) and Jacquez *k*-nearest neighbour (*k*-NN) test (Jacquez 1996). In the Knox test, critical distances are defined as subjective threshold values that dictate the spatial and temporal distance where events are deemed 'close' in both space and time. The counts of the pairs of events (i.e. two 'close' events) are then put to statistical test for clustering. Although the critical distances can be determined according to prior knowledge, the subjectivity in specifying the critical temporal and the critical spatial distance may result in significantly different results. Jacquez *k*-NN test is constructed on a statistic index of spatio-temporal nearest neighbour (STNN) pairs in which the temporal and the spatial distance between the two events is less than the *k*th spatial and the *k*th temporal nearest distance for both events (Jacquez 1996). The *k*-NN utilizes the index J_k, which is referred to as the count of STNN pairs and varies with *k*, to measure the closeness of events in the data set. The *k*-NN technique has been used in discovering crime point patterns (Cromwell *et al.* 1999, Ratcliffe 2005). It has the ability to track a sequence of identical events through time in an epidemiological context and connect the detected events to form a 'chain' of cases (Jacquez 1996). The chain begins with an index case and is spread through a contagious process to other cases. Although the detected 'chain' may reflect the sequence of contagious process, it is not a dense cluster (what we intend to discover). To identify spatio-temporal clustering patterns, Zaliapin *et al.* (2008) introduced the time–space–magnitude distance to extract aftershocks from background earthquakes. In their research, the time–space–magnitude distance between two events is defined as a function of spatial distance, time interval and magnitude difference between earthquakes. Although the time–space–magnitude distance merges the space, time and magnitude factor in one concept, the time interval should be defined according to prior knowledge and the separation of feature and noise can only be realized interactively. Wang *et al.* (2006) and Birant and Kut (2007) proposed the ST-DBSCAN algorithm to identify clusters from spatio-temporal data in the presence of noise objects. In ST-DBSCAN, clusters are formed by connecting objects whose *k*th temporal and *k*th spatial nearest distances are less than the temporal and the spatial thresholds, respectively, which are interactively estimated with the whole data. There are two defects in ST-DBSCAN. The first is that the threshold may be underestimated and consequently lead to the overestimation of clusters. The second is it cannot avoid interactive trials on the estimation of the spatial and the temporal thresholds.

In summation, existing methods either need to estimate local density in a subjective way or need to calculate specific statistics for clustering testing, which requires more prior knowledge. In this article, we propose a new approach to spatio-temporal clustering: windowed nearest neighbour (WNN) clustering method. The WNN method assumes that clusters and noise belong to different point processes. The one with a higher intensity can be deemed as clusters and the one with a lower intensity can be treated as noise. Our method is composed of the following steps. The first is to construct the sequence of temporal window width (TWW) factors, which is generated in the sequence: $w_i = \left\{ \frac{(-1)^i + 5}{2 * 2^{\text{floor}(i/2)}} \right\} (i = 1, 2 \ldots N)$, where floor() is the function that maps a real number to the next smallest integer. The TWW can thus be calculated by multiplying the time scope of data by TWW factors. In the second step, the spatio-temporal densities of feature are calculated at different TWWs; the appropriate values of TWW for the classification can be located at the local maximum values of the densities. In the third step, feature and noise are separated at the determined TWW by using expectation-maximization (EM) algorithm. The final step is to connect events into clusters according to their density connectivity in geographic–temporal space.

This article is structured as follows. Section 2 introduces basic concepts used in the WNN method. In Section 3, the algorithm of WNN is described and the estimation of parameters is discussed in detail. The algorithm is then evaluated with simulated data in Section 4. A seismic case study is presented in Section 5. Section 6 provides a summary of this article as well as directions for future research.

2. Basic concepts

2.1. Windowed kth nearest distance

Recall that the *k*th nearest distance of event p in the geographic space is defined as the distance between p and its *k*th nearest neighbour (Byers and Raftery 1998). In this article, we extend the concept to the temporal–geographic space and introduce the concept of windowed *k*th nearest (WKN) distance.

Definition 1 (spatio-temporal neighbourhood of event $p_i(x_i, t_i)$ ($A_{\Delta S,\Delta T}(p_i)$): A spatio-temporal cylinder centred at p_i with a geographic radius of ΔS and a TWW of ΔT (Figure 1).

Definition 2 (WKN neighbour of event $p_i(x_i, t_i)$ ($p_{i+k}(x_{i+k}, t_{i+k})$)): The *k*th geographic nearest neighbour within the spatio-temporal neighbourhood of $p_i(x_i, t_i)$ ($A_{\infty,\Delta T}(p_i)$).

Note that the geographic radius of $A_{\infty,\Delta T}(p_i)$ is infinity in Definition 2. We also can deduce that for k windowed nearest neighbours of $p_i(x_i, t_i)$ $(p_{i+1}(x_{i+1}, t_{i+1}), p_{i+2}(x_{i+2}, t_{i+2}), \ldots, p_{i+k}(x_{i+k}, t_{i+k}))$, $|t_{i+j} - t_i| \le \Delta T$ $(j = 1, 2, \ldots k)$ and $\|x_{i+1} - x_i\| \le \|x_{i+2} - x_i\| \le \cdots \le \|x_{i+k} - x_i\|$.

Definition 3 (WKN distance of event $p_i(x_i, t_i)$) ($d_{k,\Delta T}(p_i)$): The geographic distance between p_i and its WKN neighbour is $p_{i+k}(x_{i+k}, t_{i+k})$, that is, $d_{k,\Delta T}(p_i) = \|x_{i+k} - x_i\|$.

Definition 4 (spatio-temporal core event): A spatio-temporal core event with respect to $(\Delta S, \Delta T)$ and MinPts is the event whose neighbourhood $A_{\Delta S,\Delta T}(p_i)$ contains at least MinPts events.

Definition 5 (spatio-temporal density-connectivity): An event p is considered to be spatio-temporal density-connected to event q with respect to $(\Delta S, \Delta T)$ and MinPts if there is a collection of events $p_1, p_2, \ldots, p_n$ (with $p_1 = q$ and $p_n = p$) so that $p_{i-1} \in A_{\Delta S,\Delta T}(p_i)$ $(i = 2, 3, \ldots, n)$ and $A_{\Delta S,\Delta T}(p_i)$ $(i = 2, 3, \ldots, n - 1)$ must contain at least MinPts events.

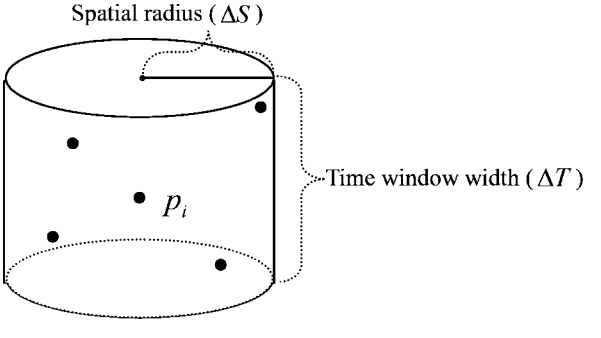

Figure 1. Spatio-temporal neighbourhood of event p_i.

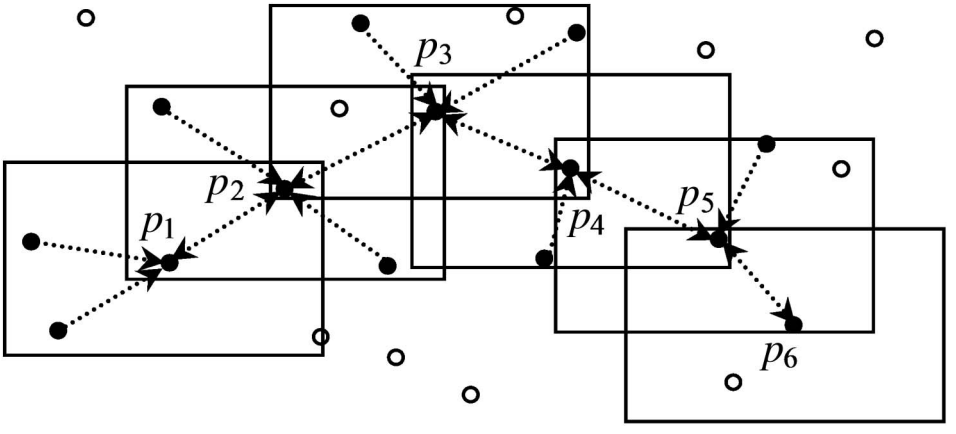

Figure 2. Spatio-temporal density connectivity. Rectangles symbolize $A_{\Delta S,\Delta T}(p_i)$ (from a horizontal perspective) with respect to ΔS, ΔT and 3; arrows symbolize the linkages between events and their core events; circles indicate the events not located in any $A_{\Delta S,\Delta T}(p_i)$.

Figure 2 shows a sequence of spatio-temporal density-connected events. p_1 is spatio-temporal density-connected to p_6 with respect to $(\Delta S, \Delta T)$ and 3 (MinPts).

2.2. *Assumption about spatio-temporal Poisson point process*

A point process P can be deemed as a homogeneous Poisson point process if the number of events (k) in any unit of $S \subseteq X$ with volume $|S|$ follows the distribution below:

$$f_{\lambda|S|}(k) = \frac{e^{-\lambda|S|}(\lambda|S|)^k}{k!} \tag{1}$$

where k is the number of events in $S(k = P(S))$, X is the support domain of P (the region in which the point process P is constrained) and λ is the intensity of the process, which is defined as the ratio between the number of events in P and $|X|$ (Byers and Raftery 1998). This says that the events of the homogeneous Poisson point process are equally likely to occur anywhere within X and do not interact with each other. Based on the concept of homogeneous Poisson point process, we give the definition of spatio-temporal homogeneous Poisson point process and ΔT intensity.

Definition 6 (spatio-temporal homogenous Poisson point process): For a homogeneous Poisson point process P, if the support domain of P is a subset of the geographic–temporal space, we then denote P as a spatio-temporal homogeneous Poisson point process (for short, we use ST Poisson process hereafter). In this article, we use intensity for a point process and density for a cluster.

Definition 7 (ΔT intensity for a ST Poisson process): ΔT intensity ($\lambda_{\Delta T}$) for a ST Poisson process is defined as

$$\lambda_{\Delta T} = \Delta T \cdot \lambda \tag{2}$$

where λ is the intensity of the ST Poisson process. Note that λ is a constant for an ST Poisson process. Given an event p_i in an ST Poisson process, $\lambda_{\Delta T}$ of $A_{\Delta S,\Delta T}(p_i)$ can be computed as

$k/(\pi\Delta S^2)$, where k is the number of events in $A_{\Delta S,\Delta T}(p_i)$, and λ of $A_{\Delta S,\Delta T}(p_i)$ can be computed as $k/(\pi \cdot \Delta S^2 \cdot \Delta T)$ or simply as $\lambda_{\Delta T}/\Delta T$.

Assumption 1 If a spatio-temporal event set ($B\ \{p_i(x_i, t_i)\ \ i = 1, 2, ...N\}$) contains clusters and noise, then clusters and noise can be treated as two distinctive ST Poisson point processes $\left(P_j\ (j = 1, 2)\right)$ with different intensities, where x_i is the geographic location of the event p_i and t_i is the time attribute of the event p_i. Clusters are distributed at the higher intensity and noise is distributed at the lower intensity.

Note that here we exclude the situation where the event set includes several clusters with different densities. (In fact, our method can also handle data containing multiple processes [or clusters], which will be discussed in the latter part of this article.) The assumption enables us to use the intensity (density) to differentiate between clustered events and noise. The definition of ST Poisson point process shows that events in a homogeneous Poisson point process are independently and uniformly distributed over X. As a result, for an ST Poisson process $P_\lambda\{p_i(x_i, t_i)\}$, the spatial point process ($\{x_i\}$) and the time point process ($\{t_i\}$) can be viewed as two independent homogeneous Poisson point processes, and the geographic location (x_i) is independent of the time attribute (t_i).

2.3. Probability density function of WKN distance

For a given event p_i in an ST Poisson process, the probability density function (pdf) of its WKN distance ($d_{\Delta T,k}$) can be acquired through seeking the probability function of including 0, 1, 2, . . ., $k-1$ events within its spatio-temporal neighbourhood ($A_{\Delta S,\Delta T}(p_i)$):

$$P(d_{\Delta T,k} \geq x) = \sum_{m=0}^{k-1} \frac{e^{-\lambda_{\Delta T}\pi x^2}(\lambda_{\Delta T}\pi x^2)^m}{m!} = 1 - F_{d_{\Delta T,k}}(x) \tag{3}$$

where $F_{d_{\Delta T,k}}(x)$ is the cumulative distribution function of $d_{\Delta T,k}$ and $\lambda_{\Delta T}$ is the ΔT intensity of $A_{\Delta S,\Delta T}(p_i)$. If $d_{k,\Delta T}$ is larger than x, there must be 0 or 1 or 2 . . . $k-1$ events within $A_{\Delta S,\Delta T}(p_i)$ and its pdf $\left(f_{d_{\Delta T,k}}(x)\right)$ is the derivative of $F_{d_{\Delta T,k}}(x)$:

$$f_{d_{\Delta T,k}}(x) = \frac{dF_{d_{\Delta T,k}}(x)}{dx} = \frac{e^{-\lambda_{\Delta T}\pi x^2} 2(\lambda_{\Delta T}\pi)^k x^{2k-1}}{(k-1)!} \tag{4}$$

where $\lambda_{\Delta T}$ and k are the same as those in Equation (3). The pdf can be treated as a mixture pdf of gamma according to the definition of pdf of gamma, that is, $Y \sim \Gamma(k, \lambda_{\Delta T}\pi)$, where $Y = x^2$ (Byers and Raftery 1998).

3. Theory of windowed nearest neighbour

3.1. WNN method

The WNN cluster method is based on the concept of the WKN distance and is composed of two stages. The first is to separate clustered events (feature process) from noise; the second is to form distinctive clusters from the clustered events. In the first stage, the local spatio-temporal density can be employed to differentiate clustered events from noise based on Assumption 1. So the determination of the threshold of local spatio-temporal density is the key to the first stage. In the WNN method, we use the WKN distance instead of spatio-temporal density. Once the threshold of the WKN distance ($D_{\Delta T,k}$) is determined, that threshold and the parameters (i.e. k and ΔT) associated with it define a cylinder. The density of the

cylinder is $k/(\pi D^2_{\Delta T,k}\Delta T)$. Thus, $D_{\Delta T,k}$ can be treated as a replacement of the threshold of density. As a result, events are classified as feature if their WKN distances are no larger than $D_{\Delta T,k}$ while events are classified as noise if their WKN distances are larger than $D_{\Delta T,k}$. $D_{\Delta T,k}$ can be estimated by using EM algorithm (readers are referred to Appendix for more details).

In the second stage of WNN, the cylinder defined by $D_{\Delta T,k}$, ΔT and k is then used for connecting events into clusters by linking the spatio-temporal density-connected events (which was defined in Section 2). Interested readers may refer to Ester *et al.* (1996), Wang *et al.* (2006) and Birant and Kut (2007) for details. Here, we name the cylinder as the spatio-temporal density-connectivity unit. Below we first describe the WNN algorithm, and then discuss the parameter choice problem.

Algorithm: ***WNNCluster*** (Input: *Data*, *k*, *DeltaT*)

```
    WKNDistance = CalculateWKNDistance(Data, DeltaT, k);
    [DeltaS, Flamda, M] = MixtureDecomByEM(WKNDistance, k);
    FeatureSet = Data(M>=0.5); // Events are classified as feature if M ≥ 0.5.
  Noise = Data(M<0.5); // Events are classified as noise if M < 0.5.
    ClusterId := 1;
    for i = 1 to FeatureSet.size() do
      Event := FeatureSet.get(i); // Get each event from FeatureSet
      if Event.Id == UNCLASSIFIED THEN
        if ExpandCluster(FeatureSet, Event, ClusterId, DeltaT, DeltaS, k)
          ClusterId := ClusterId + 1;
        end if
      end if
    end for
  end algorithm;
  function CalculateWKNDistance(Data, DeltaT, k): Array
    for i=1 to Data.size()
      Pi= Data.get(i);
      for j=1 to Data.size()
        if j<>i
          Pj= Data.get(j);
          if Pj.t<Pi.t + DeltaT /2 & Pj.t>Pi.t –DeltaT /2
            DataT.Add(Pj);
          end if
        end if
        WKNDistance(i) = CalculateDis(Pi, DataT, k);
      end for
    end for
    DataT.clear();
    return WKNDistance;
```

In function *WNNCluster*(), function *CalculateWKNDistance*() returns the WKN distances of all events calculated at the specific *DeltaT*(ΔT). Function *MixtureDecomByEM*() estimates the threshold of the WKN distance (*SThreshold*) by EM algorithm and returns the ΔT intensity (*Flamda*) of the feature and the fuzzy membership values (M(i)) of events belonging to feature at the specific *DeltaT* (for computational details, please refer to the Appendix). Function *ExpandCluster*() is to form clusters by linking spatio-temporal density-connected events (with respect to *DeltaT*, *DaltaS* and *k*) into clusters.

Nevertheless, two parameters (i.e. ΔT and k) need to be tuned before determining the threshold of the WKN distance. To analyse the sensitivity of WNN to the parameters, in Section 3.2 we first use simulated data to illustrate the effect of ΔT and then give the solution regarding how to determine ΔT. The effect and choice of k are discussed in Section 3.3.

3.2. *Determination of* ΔT

Figure 3 displays a simulated spatio-temporal data that contains a 'cubic' feature and noise. We then use different values of ΔT, acquired by multiplying the time scope of data by varying TWW factors sampled at the sequence $w_i = \left\{\frac{(-1)^i+5}{2*2^{\text{floor}(i/2)}}\right\}$ $(i = 1, 2, \ldots, N)$ in (0 2] (i.e. 2, 3/2, 1, 3/4, 1/2, 3/8, 1/4, . . .), to classify the data. The results are found to be varying with the TWW factor; see Table 1. In other words, ΔT has a significant influence on the classification in terms of the number of false positives and cluster number. In the following text, we first try to explain the results based on the analysis of the relationship between the density of identified feature, the number of false positives and ΔT, and then provide the algorithm for estimating ΔT.

3.2.1. *Relationship between density of identified feature and* ΔT

The curve of the density of identified feature versus TWW factor (k = 10) is shown in Figure 4 and that of the number of false positives versus TWW factor is shown in Figure 5. Note that the edge correction was made by simulating data with the same density as noise in the buffered space, whose buffered distances are halves of the scopes of the data in *X*-, *Y*- and *T*-directions, respectively. Even though, as TWW factor decreases to an extremely small value, events in the data may fail to find its WKN neighbour. If events in a data set, which cannot find their WKN neighbours, are more than 30% of the total count, the value of TWW factor will be highlighted (namely, symbols are indicated by circles in Figure 4 and thereafter). From Table 1 and Figures 4 and 5, we find that (1) there exists one peak (which is located in the interval of TWW factor between 3/64 and 3/16) on the curve of density of

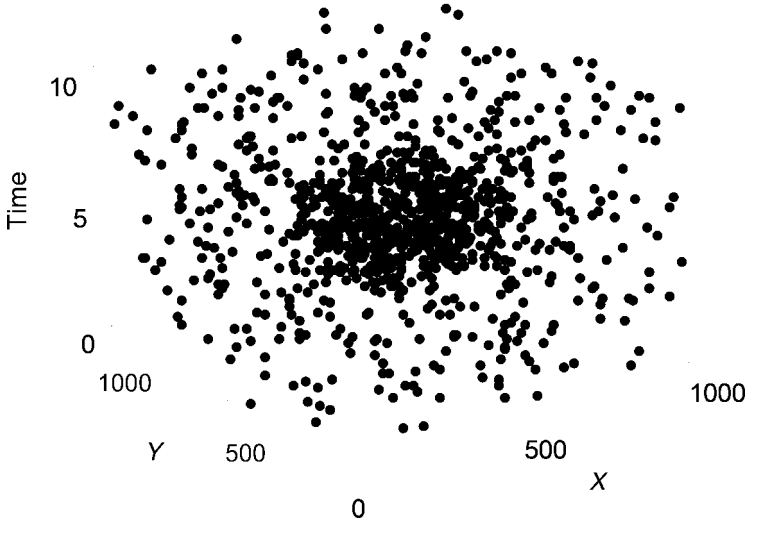

Figure 3. Simulated spatio-temporal data.

Table 1. Results of simulated data at different values of TWW factor.

Time window factor		1/512	3/1024	1/256	3/512	1/128	3/256	1/64	3/128	1/32	3/64	1/16
Threshold ($k = 10$)		Null	474.6	468.4	434.2	397.3	363.7	335.2	281.2	236.5	183.3	156.7
Number of false positives and clusters	$k = 10$	432 (0)	409 (8)	356 (18)	213 (19)	113 (9)	66 (3)	41 (1)	31 (1)	23 (1)	22 (1)	21 (1)
	$k = 3$	192 (58)	143 (51)	106 (31)	95 (24)	65 (15)	64 (12)	44 (9)	50 (10)	46 (8)	40 (8)	39 (6)
	$k = 20$	Null	433 (0)	433 (0)	433 (0)	394 (9)	193 (9)	108 (1)	47 (1)	36 (1)	29 (1)	27 (1)
Density (10^{-4})	$k = 10$	1.93	3.08	3.38	3.85	4.01	4.34	4.89	5.34	5.65	6.17	6.30
	$k = 3$	6.15	5.05	5.68	5.95	6.19	6.85	6.87	6.93	7.05	7.64	7.28
	$k = 20$	No data	No data	1.39	1.93	1.78	2.58	3.28	4.20	4.67	5.14	5.54
Time window factor		3/32	1/8	3/16	1/4	3/8	1/2	3/4	1	3/2	2	
Threshold ($k = 10$)		128.5	111.0	91.5	80.1	68.0	61.9	56.5	53.8	50.7	49.9	
Number of false positives and clusters	$k = 10$	26 (1)	29 (1)	24 (1)	35 (1)	34 (1)	46 (1)	65 (2)	83(2)	90 (3)	92 (3)	
	$k = 3$	42 (7)	53 (9)	60 (14)	55 (14)	62 (10)	66 (6)	91 (10)	109 (13)	131 (16)	151 (19)	
	$k = 20$	27 (1)	30 (1)	29 (1)	36 (1)	42 (1)	48 (1)	65 (1)	81 (1)	90 (1)	91 (1)	
Density (10^{-4})	$k = 10$	6.26	6.23	6.14	5.79	5. 29	4.63	3.38	2.67	1.87	1.41	
	$k = 3$	6.94	6.54	6.47	6.23	5.71	4.85	3.58	2.88	2.07	1.58	
	$k = 20$	5.79	5.75	5.77	5.46	5.05	4.40	3.30	2.64	1.85	1.40	

Note: Numbers in parentheses are cluster numbers.

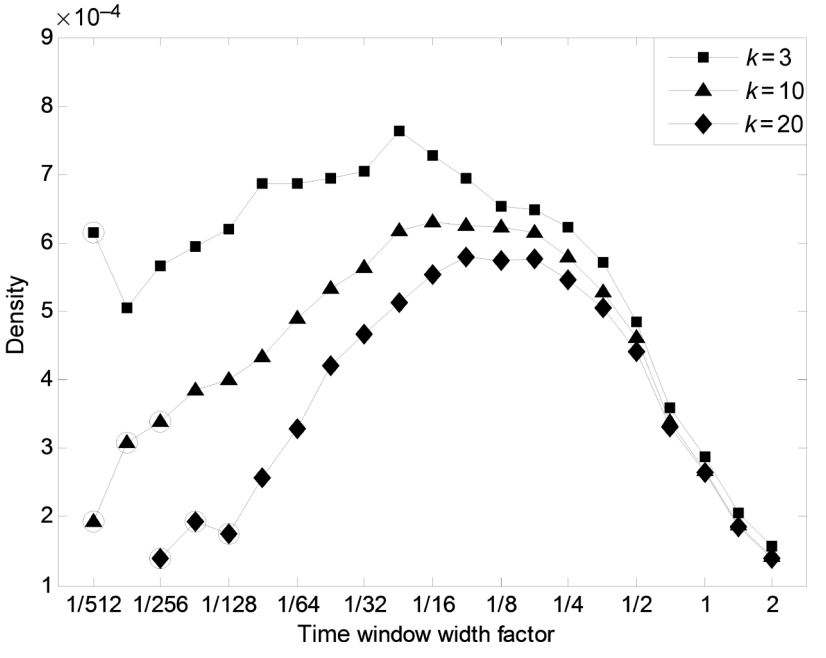

Figure 4. Density of identified feature varied with TWW factor (for $k = 10$, the local maximum value of density is generated when TWW factor = 1/16; for $k = 3$, the local maximum value of density is generated when TWW factor = 3/64; for $k = 20$, the local maximum value of density is generated when TWW factor = 3/32; symbols enclosed with circles indicate results in which more than 30% events cannot find their WKN neighbours hereafter).

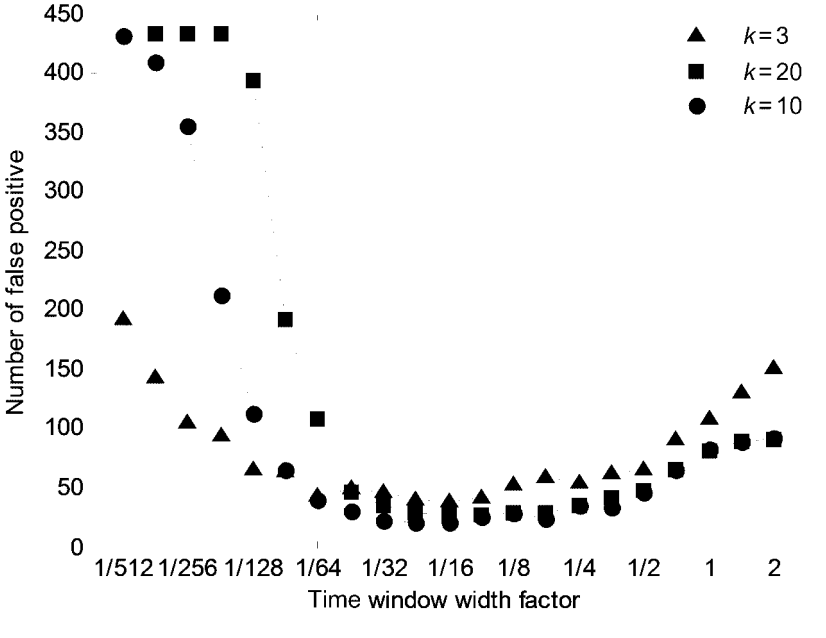

Figure 5. Number of false positives varied with TWW factor and k.

identified feature and (2) the local maximum value of the density of identified feature and the minimum number of false positives are both generated as TWW factor = 1/16.

From the analysis, we conclude that the highest density of identified feature might indicate a small number of false positives, essentially, the value of ΔT, which produces the highest density, can be adopted for the classification.

To explain the dependence of the number of false positives on ΔT, we chose classifications generated as TWW factor = 2, 1/16 and 3/128 (i.e. $\Delta T = 20$, 5/8 and 15/64) and $k = 10$. Figure 6 sketches the rectangle feature of Figure 3 and different spatio-temporal density-connectivity units with the heights of ΔTs mentioned above. Figure 7 shows the histograms of $d_{\Delta T,k}$ of events in Figure 3 and the classifications generated at different ΔTs. When ΔT is longer than the time scope of feature (e.g. Unit A in Figure 6, $\Delta T = 20$), k windowed nearest neighbours of a given feature event p_i inevitably contain noise events. Furthermore, noise events over and below a feature cannot be separated from the feature, thereby leading to more false positives (Figure 7b). This can also be validated by the fitted curve of the histogram of the WKN distance (Figure 7a), in which the proportion of feature is larger than the theoretical (the left component) and that of noise is smaller than the theoretical (the right component). Because of the existence of many noise events in the neighbourhoods of feature events, ΔT intensity ($\lambda_{\Delta T}$) of identified feature events is underestimated. Consequently, the spatio-temporal density of identified feature, which equals $\lambda_{\Delta T}/\Delta T$ (recall Definition 7), is lowered (1.41×10^{-4}) and the feature is then overestimated in terms of event number and cluster number (92 false positives and 3 clusters are produced).

As ΔT decreases (Unit B in Figure 6, $\Delta T = 5/8$), fewer noise events are included in k windowed nearest neighbours of feature events. Those near the centre of feature may even exclude noise events from their k windowed nearest neighbours. The 'purified' k windowed nearest neighbours may reduce the probability of events being misclassified, which is also supported by the more distinctive mixture histogram in Figure 7c (compared with that in

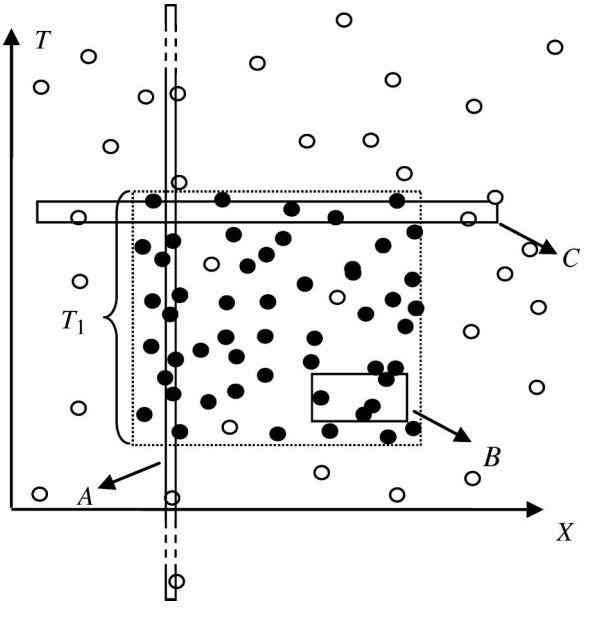

Figure 6. Spatio-temporal density-connectivity unit with varying ΔT (dots indicate the feature events and circles indicate noise; A, B and C are spatio-temporal density-connectivity units; their TWW factors are 2, 1/16 and 3/128 respectively).

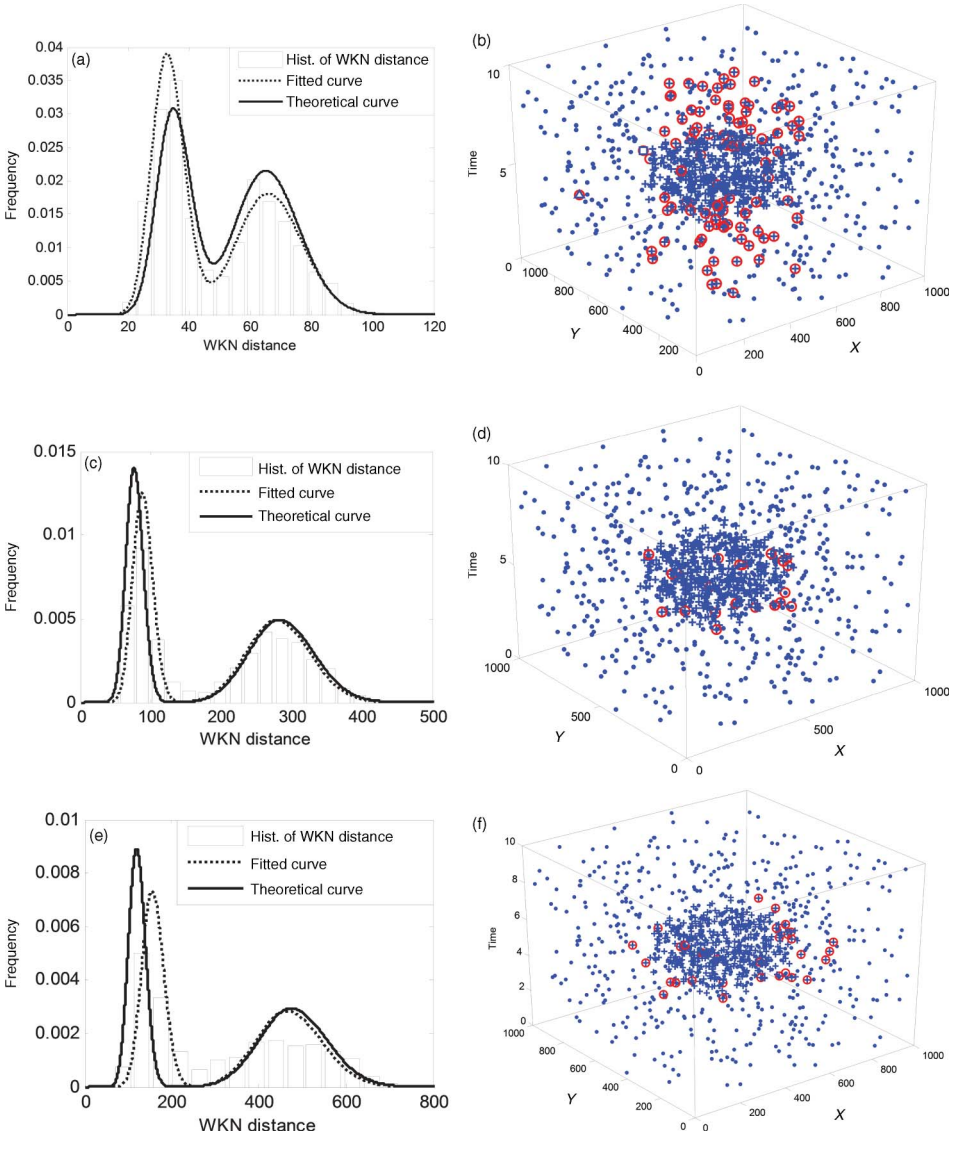

Figure 7. Results of simulated data using WNN: (a) mixture histogram of the WKN distance when TWW factor = 2; (b) classification when TWW factor = 2 ($\Delta T = 20$); (c) mixture histogram of the WKN distance when TWW factor = 1/16; (d) classification when TWW factor = 1/16 ($\Delta T = 5/8$); (e) mixture histogram of the WKN distance when TWW factor = 3/128; (f) classification when TWW factor = 3/128 ($\Delta T = 15/64$).

Figure 7a). Because events in neighbourhoods ($A_{d_{\Delta T,k},\Delta T}(p_i)$) of a given feature event become purer as ΔT decreases, the density of identified feature will reach a higher value (6.3×10^{-4}). At the same time, false positives are reduced and no false cluster is generated (21 false positives; see Table 1 and Figure 7d).

As ΔT decreases to a small value (e.g. Unit C in Figure 6, $\Delta T = 15/64$), the WKN distance of a feature event is forced to be larger than the theoretical event because of the difficulty in finding enough nearest neighbours of same process within such a small ΔT. The

histogram of feature (the left component) shows significant right-bias (Figure 7e) compared with that in Figure 7c. This will lead to more false positives (31 false positives; Figure 7f). Because of the existence of noise events in the neighbourhood of a given feature event ($A_{d_{\Delta T,k},\Delta T}(p_i)$), the density of feature is underestimated (5.34×10^{-4}; Table 1) and the number of false positives increases accordingly (31 false positives).

To sum up, densities calculated at differing ΔTs may indicate different numbers of false positives. As ΔT decreases to a value at which most feature events could find their k windowed nearest feature neighbours, the density of identified feature may reach its (local) maximum value(s), and at the same time the number of false positives may reach a low value (in this case, Unit B in Figure 7). Note that local maximum values can also indicate appropriate values of ΔT when clusters with different densities and predominant dimensions coexist, which will be discussed in the results of other simulated data sets (see Section 4).

3.2.2. *Algorithm for estimating* ΔT

Based on the analysis of the relationship between the density of identified features and ΔT, we give the algorithm for estimating ΔT:

Function: [*DeltaT*] = *DetermineDaltaT*(*Data*, *k*)
let *TimeWindowWidthFactor* = [1/512, 3/1024, 1/256, 3/512, 1/128, 3/256, 1/64, 3/128, 1/32, 3/64, 1/16, 3/32, 1/8, 3/16, 1/4, 3/8, 1/2, 3/4, 1, 3/2, 2];
for *i* = 1 **to Length**(*TimeWindowWidthFactor*)
  TimeWindowWidth = *TimeWindowWidthFactor*(*i*)**Data.GetTScope*(); //(1)
  WKNDistance = *CalculateWKNDistance*(*Data*, *TimeWindowWidth*, *k*);
  [*DeltaS*, *Flamda*, *M*] = *MixtureDecomByEM*(*WKNDistance*, *k*);
  Intensity(*i*) = *Flamda* / *TimeWindowWidth*; //(2)
end for
  [*DeltaT*] = *LocalMaxIntensity*(*Intensity*); //(3)
end algorithm //(4)

Regarding the algorithm, some points should be noted:

(1) *TimeWindowWidth* is the width of time window; function *Data.GetTScope*() returns the time scope of the spatio-temporal data.
(2) The spatio-temporal intensity (density) of feature is estimated by dividing *Flamda* (ΔT intensity) by *TimeWindowWidth*.
(3) Function *LocalMaxIntensity*() returns the appropriate value(s) of ΔT, at which the local maximum values of density of feature are generated. In *LocalMaxIntensity*(), the local maximum values of density are determined in the sequence of densities acquired over varying TWW factors.
(4) Because parameters (i.e. *M*, *Flamda* and *DeltaS*) associated with the estimated ΔT have been determined in Function *DetermineDaltaT*(), the same steps in *WNNCluster*() can be omitted.

3.3. *Choice of k*

We then examine the influence of k on the classification. Besides $k = 10$, Figure 4 shows the plot of the density of identified features versus TWW factor when $k = 3$ and $k = 20$. The numbers of false positives and clusters can be found in Table 1 while the curves of the number of false positives are displayed in Figure 5. According to the comparison of the

numbers of false positives between $k = 3$ and $k = 10$ (Figure 5 and Table 1), results for $k = 3$ show more false positives than those for $k = 10$ when TWW factor exceeds 1/64. More importantly, the smallest number of false positives for $k = 3$ is larger than that for $k = 10$, both of which were generated as TWW factor = 1/16. In addition, the number of clusters is overestimated at all values of TWW factor as $k = 3$. Differing from those for $k = 3$, results for $k = 20$ show more false positives than those for $k = 10$ almost at all values of TWW factor except when it is larger than 1/2 (Table 1 and Figure 5). The smallest number of false positives for $k = 20$ is larger than that for $k = 10$. Nevertheless, results for $k = 20$ show wrong cluster number only when TWW factor is less than 1/64. The comparison of results between $k = 3$, $k = 10$ and $k = 20$ shows that the most appropriate value for k is 10.

The determination of k for the NN method has been discussed in Pei *et al.* (2007). Here we only recapitulate the result that is also applicable to the scenario of WNN. As k is small, the histogram of the WKN distance is not clearly bimodal (Figure 8a), which leads to many false positives being produced. As k is too large, because the WKN distances of events (either noise or feature) near the border between feature and noise may significantly become smaller (for noise) or larger (for feature) than those far away from the border, the mixture histogram of the WKN distance deviates from the theoretical one (Figure 8b). The feature will be overestimated while the noise will be underestimated. Therefore, the algorithm may not produce a good result in this case. In conclusion, if one cares about the total number of false positives we suggest $k = 6$–12; if one cares only about the number of false positives of feature (i.e. the number of feature events that have been classified as noise), for example, in the case of predicting the susceptible area of natural disasters, a larger value of k will be appropriate. In this article, we are concerned more about the number of false positives and set k to a medium large value in the examples of simulated data and the case study.

4. Results of other simulated data

4.1. *Classification results of simulated data*

To evaluate the WNN method, we simulated five data sets (Figure 9). The first data set consists of five clusters and noise; the second consists of two 'L-shaped' clusters and noise; the third contains two clusters – one is contained in a cuboid and the other is limited in a thin plane extended in *XY* dimension; the fourth contains two clusters (with the same number of events) – one is in a high density, and the other is in a low density and prolongs along the

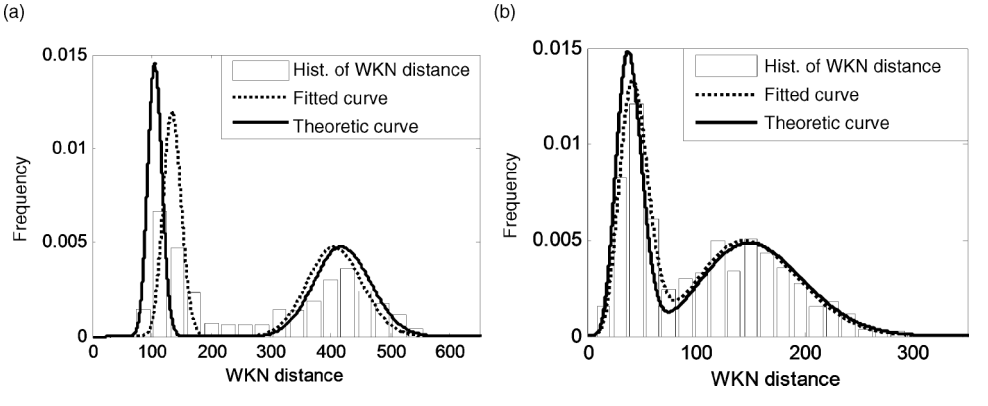

Figure 8. Histograms generated at different values of k (TWW factor = 1/16): (a) histogram of the WKN distance as $k = 3$; (b) histogram of the WKN distance as $k = 20$.

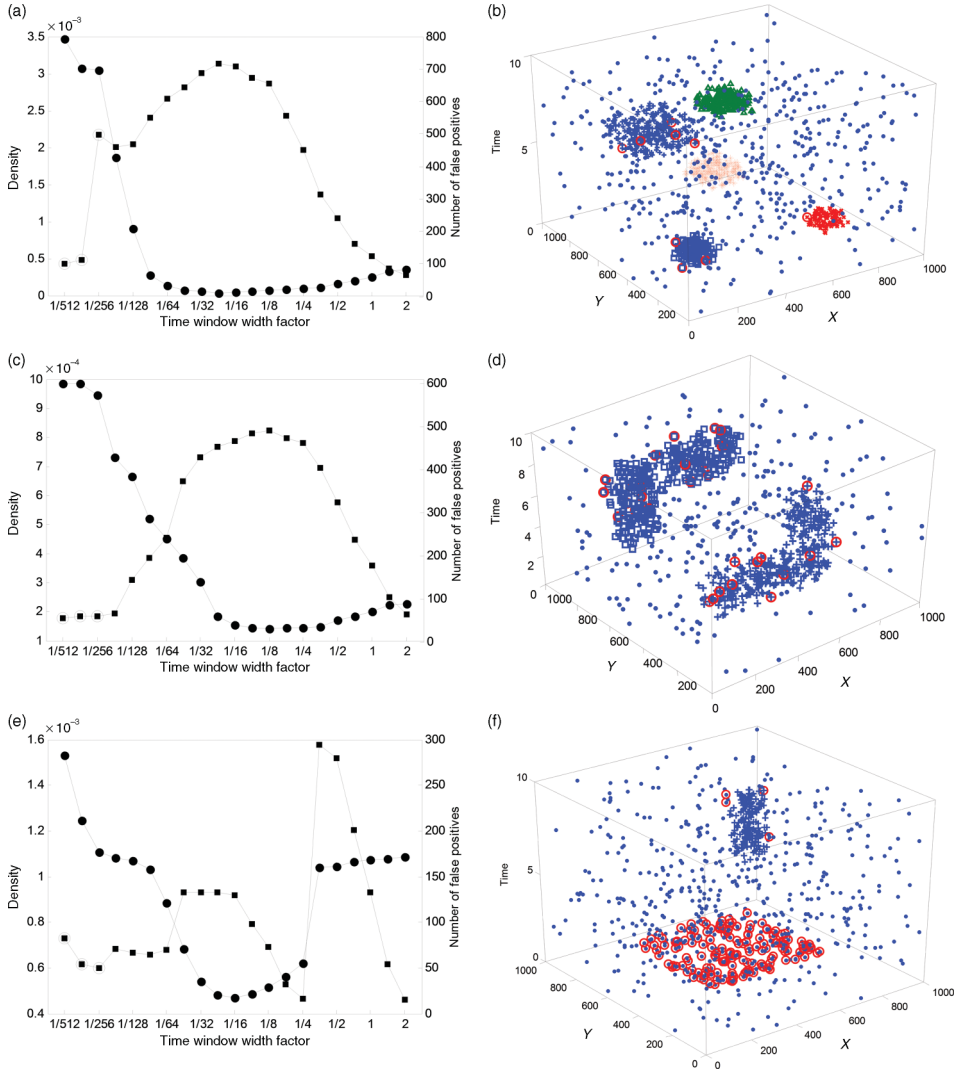

Figure 9. Results of simulated data: (a) densities and numbers of false positives of the first data set; (b) classification (TWW factor = 3/64, false positive number = 10); (c) densities and numbers of false positives of the second data set; (d) classification (TWW factor = 1/8, false positive number = 30); (e) densities and numbers of false positives of the third data set; (f) classification (TWW factor = 3/8, false positive number = 160); (g) classification (TWW factor = 3/64, false positive number = 20); (h) classification (TWW factor = 3/512, false positive number = 170); (i) densities and numbers of false positives of the fourth data set; (j) classification (TWW factor = 3/16, false positive number = 55); (k) classification (TWW factor = 3/128, false positive number = 150); (l) densities and numbers of false positives of the fifth data set; (m) classification (TWW factor = 1/16, false positive number = 17); (n) densities and numbers of false positives of the fifth data set (identifying clusters in the rest of data in which the denser cluster has been taken out); (o) classification (TWW factor = 3/8, false positive number = 39); (p) final classification of the fifth data set. In (a), (c), (e), (i), (l) and (n), squares indicate spatio-temporal densities of identified features, dots indicate numbers of false positives; in the rest of the panels, noise is indicated by dots and false positives are highlighted by circles.

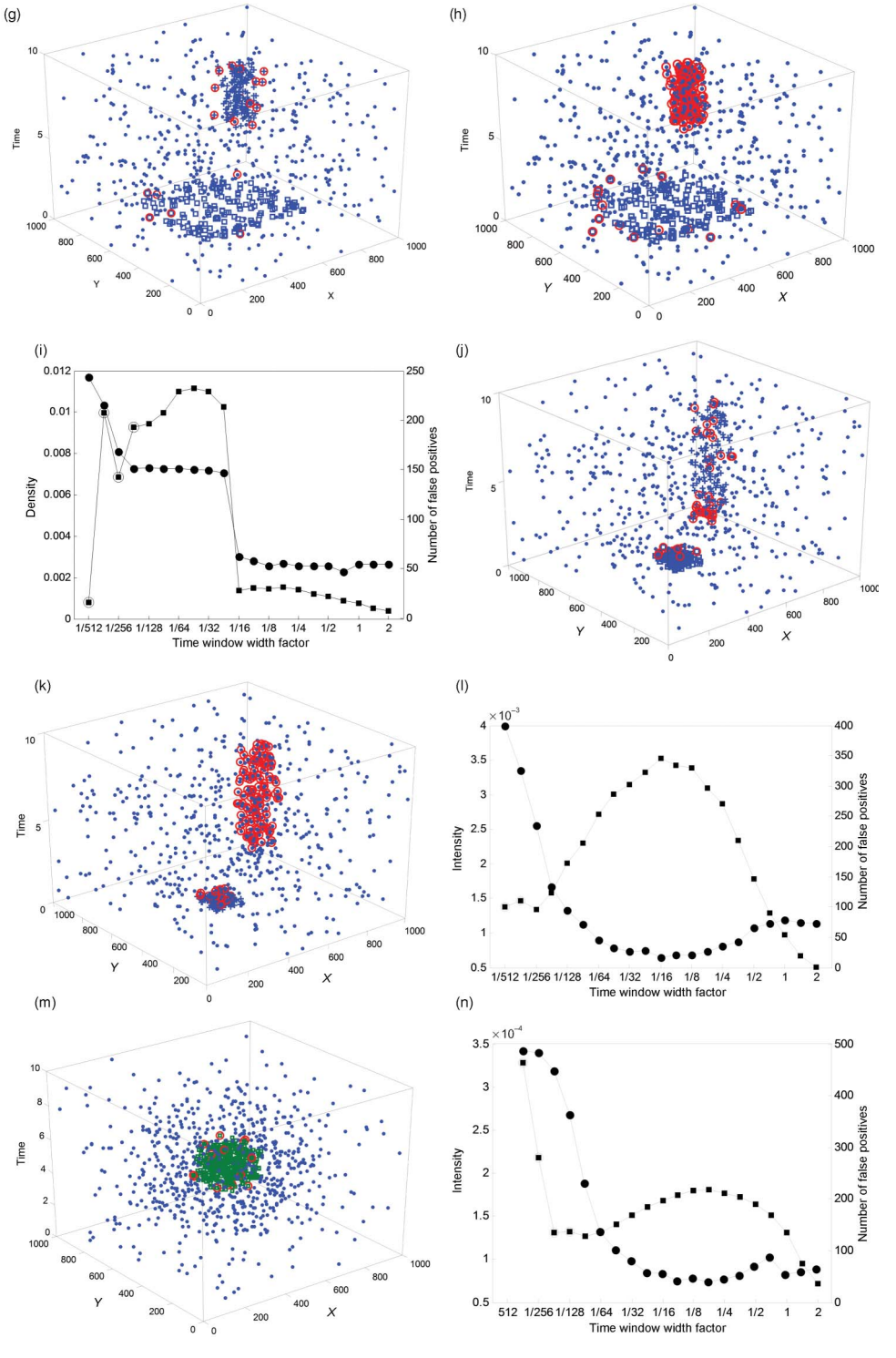

Figure 9. (*Continued*)

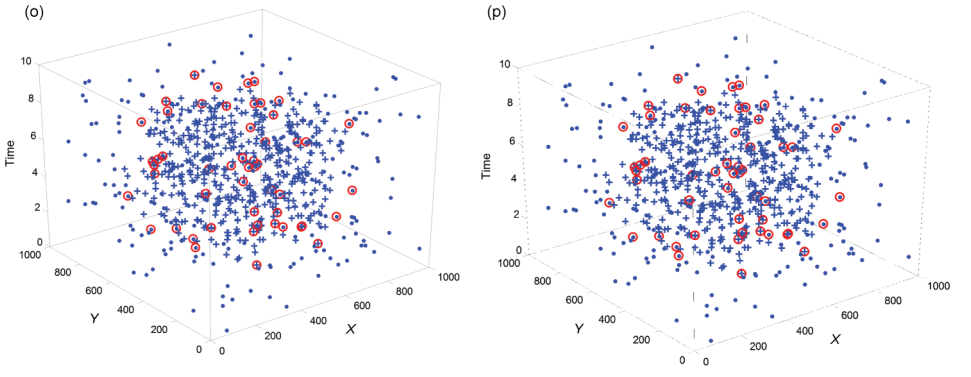

Figure 9. (*Continued*)

T-axis; the fifth contains two cubic clusters – the small and dense one is located inside the big and sparse one. Note that the fourth and the fifth data sets contain clusters of different densities.

Figure 9 displays classification results generated by WNN. Regarding the first two data sets, we found the densities reach their maximum values when TWW factor = 3/64 and 1/8, respectively (Figure 9a and c). The curves of the number of false positives also show that when TWW factors mentioned above were adopted, the smallest numbers of false positives (12, 33) were produced for the first and the second data set. The classifications, displayed in Figure 9b and d, show that the clusters are clearly revealed.

The density plot of the third data set is shown in Figure 9e; local maximum values can be located as TWW factor = 3/8, 3/64 and 3/512. Classifications acquired at these values are displayed in Figure 9f, g and h, in which the cubic cluster, both clusters and the plane cluster were extracted in turns. Although the two clusters are in the same density, they show different predominant dimensions. As a result, the one in the cuboid was identified if TWW factor was large (the reason is that events in the cuboid have shorter WKN distances than those in the plane under such a TWW, and they were classified as feature while the rest of events were classified as noise); the one in the plane was identified if TWW factor was small (the reason is that events in the plane have shorter WKN distances than those in the cuboid under such a small TWW, and the events in the plane were classified as 'feature' while the rest of events were classified as 'noise'); both were identified when TWW factor was set to a moderate value.

Two local maximum values (generated when TWW factor = 3/16 and 3/128) can be recognized from the curve of the density of the fourth data set (Figure 9i). When TWW factor = 3/16, two clusters were identified (Figure 9j); as TWW factor decreased to 3/128, only the denser cluster was extracted (Figure 9k). The reason why clusters with different densities can be identified as TWW factor = 3/16 is that events in the cluster of lower density may have the same length of the WKN distances on average as those in the clusters of higher density under such TWW. Hence, when TWW is large enough, noise can be separated from clusters, and clusters with different densities can be identified simultaneously; when TWW is small, only the cluster with the higher density can be extracted.

For the fifth data set, the curve of density shows one maximum value (Figure 9l). The initial classification only identified the denser cluster (with 17 false positives generated) and the other cluster was still hidden in noise (Figure 9m). In the next step, we treat noise and the hidden cluster as a new data set and run our algorithm again. The curve of density is displayed in

Figure 9n. The other cluster was identified as TWW = 3/8 (at which the maximum value of density was generated); the number of false positives was 39 (Figure 9o). The final result, which is the combination of these two classifications, is shown in Figure 9p.

All results show that local maximum values of density of identified features may indicate appropriate values of ΔT for classification. In particular, those of the third and the fourth data sets show that a large TWW may enhance the capability for discerning clusters extended along *T*-axis (even with a low density) but may neglect those condensed in a narrow time interval (even with a high density), whereas a small TWW may have the opposite impact. Results of the fifth data set show that the clusters with different densities can be identified stepwise by applying our algorithm several times. To sum up, for clusters with different densities and various predominant dimensions, we can try our algorithm with different TWWs (which is associated with the local maximal values of the density of identified feature) first; if it does not work, we can apply the stepwise WNN strategy.

4.2. *Comparison between WNN and ST-DBSCAN*

We used the second data set as an example to compare the efficiency on cluster identification between WNN and ST-DBSCAN. Figure 10a shows the spatial *k*-distance plot of the second data set while Figure 10b shows the temporal *k*-distance plot (*k*-distance plot displays the *k*th spatial/temporal nearest distance of each event lined up in an ascending order). In Figure 10a and b, asterisks indicate the threshold of the *k*th spatial nearest distance and that of the *k*th temporal nearest distance (ΔS and ΔT), respectively, which are located at the first valleys of the plots, as described in Birant and Kut (2007). The ST-DBSCAN was then conducted with

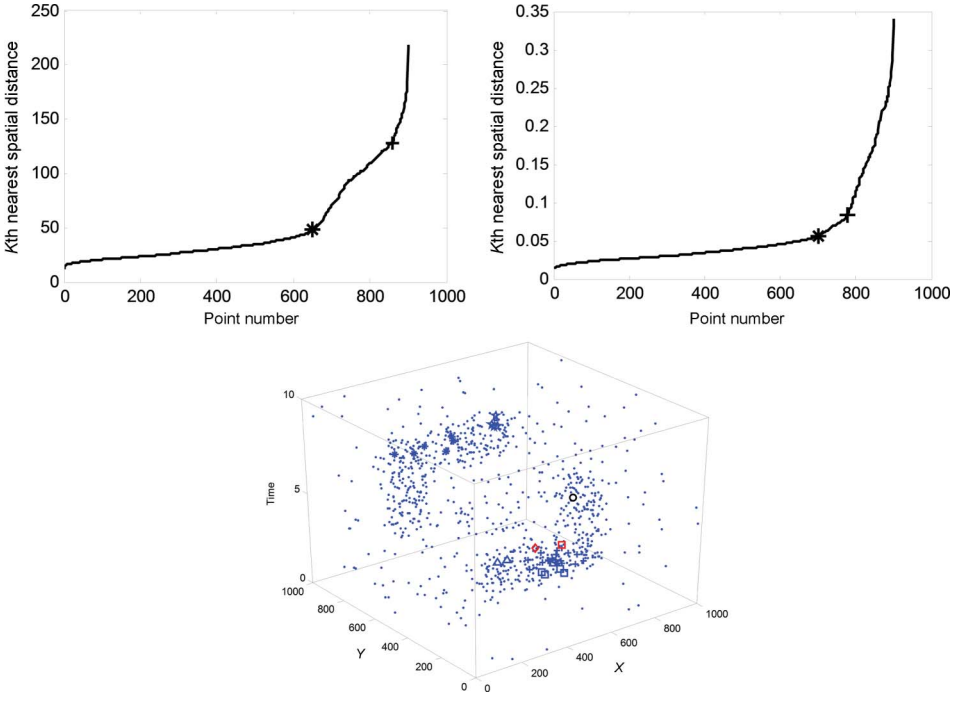

Figure 10. Classification with ST-DBSCAN: (a) spatial thresholds determined by visual trials; (b) temporal thresholds determined by visual trials; (c) classification with thresholds determined by visual trials (the number of false positive is 547).

ΔS and ΔT. However, no clusters can be found. Because the *k*-distance plots are calculated at the globe scale, the identified ΔS and ΔT are too small to discern any clusters. We then increase ΔS and ΔT, which are indicated by crosses located at valleys in Figure 10a and b, and the classification result is displayed in Figure 10c. Seven clusters were identified and 547 events were misclassified. Furthermore, similar results were also found with other data sets of Figure 9. (As we are limited by article length, we omit the details of the comparison.) As a result, we have to admit that the interactive trial may fail to produce reasonable classifications.

4.3. Complexity of WNN method

The complexity of the WNN method is decided by three factors. The first is the computation of the WKN distance of each event, whose computational complexity is $O(N^2 + N*k)$. The second is the decomposition of the mixture pdf of the WKN distance, whose computational complexity is O(*N***N*time**N*window). The third is the forming of clusters, whose computational complexity is O(*N**log(*N*)) (Ester *et al.* 1996). Here, *N* is the number of events in the data set, *Ntime* is the iteration times that the EM algorithm needs to run and *Nwindow* is the number of TWW factors that should be tried in WNN. Please note that *Nwindow* could be influenced by the data size (*N*). In detail, a large *N* may increase *Nwindow* (which means an event can find its WKN neighbour even at a very small TWW factor if the *N* events are not constrained in a limited spatial scope) while a small *N* may reduce *Nwindow* (which means an event may not find its WKN neighbour even at a relatively large TWW factor). Nevertheless, the explicit relation between *Nwindow* and *N* cannot be easily determined. Totally, the complexity of WNN is O(*N*(*N* + *k* + *Ntime***Nwindow* + log(*N*))). As a result, *N* is the key factor that decides the run time of WNN.

5. A case study: determining spatio-temporal clusters of earthquakes

5.1. Clusters of earthquakes

Clustered earthquakes are usually perceived as foreshocks (if strong earthquakes occur after them) or aftershocks (if strong earthquakes occur before them) of strong earthquakes. As a result, the detection of clustered earthquakes may help to predict strong earthquakes or understand the trend and the mechanism of strong earthquakes (Wu *et al.* 1990, Chen *et al.* 1999, Ripepe *et al.* 2000).

As we know, clustered earthquakes are difficult to discover because of the interference of background earthquakes. Nevertheless, clustered earthquakes differ from background earthquakes because clustered earthquakes are not only spatially clustered but also temporally clustered while background earthquakes are referred to as small earthquakes that release in a low stable rate (time) and intensity (space), which simultaneously occur and overlap with the clustered earthquakes (Wyss and Toya 2000, Pei *et al.* 2003). In this regard, background earthquakes and clustered earthquakes can be deemed as two ST Poisson processes with different intensities. The separation of these two types of earthquakes can be used for the evaluation of the WNN method.

5.2. Study area and seismic data

The study area is located between 101° and 106° E and 29° and 34° N (southwestern China) and is one of those areas with the most intensive seismicity in China (Figure 11). Twenty-five devastating earthquakes ($M \geq 6.0$) occurred in this area between 1970 and 2008, including

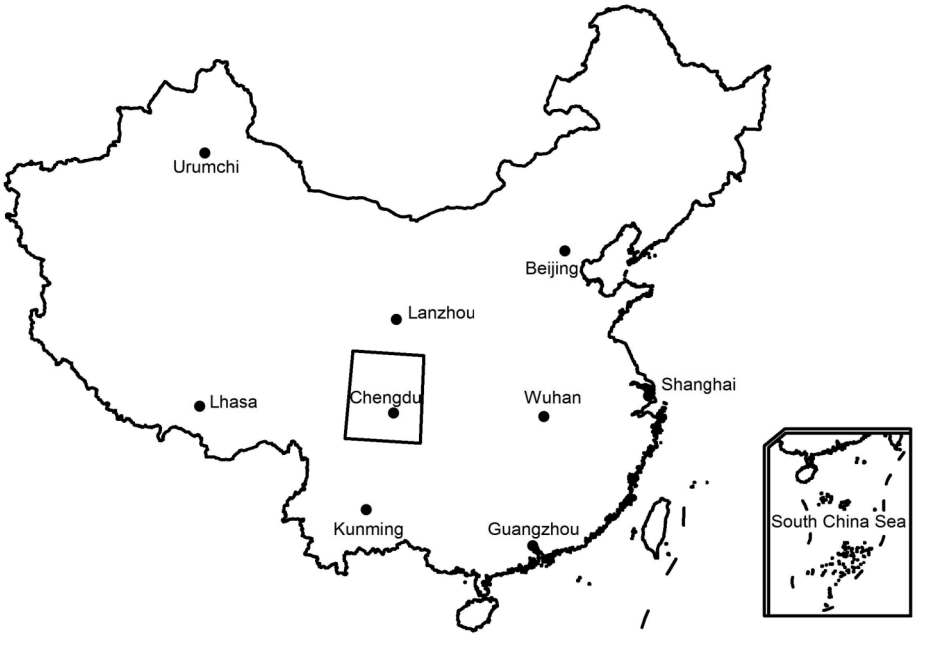

Figure 11. Location of study area.

the Wenchuan earthquake that occurred at 103.0° E, 31.3° N on 12 May 2008, with the magnitude measured as 8.0 (China Seismograph Network Data Management Center 2009). The catalogue data used for the case study are from Feng and Huang (1980, 1989). The selected earthquakes are from 15 January 1975 to 15 August 1976 and larger than 1.5 (M). Thus, 484 epicentres are obtained altogether.

5.3. Result of detection of clustered earthquakes

The WNN method was then applied to the seismic data by setting $k = 8$. The curve of density of clustered earthquakes (Figure 12) displays two platforms (i.e. TWW factor = 1/128–3/256, 1/16–2). Two local maximum values were obtained as TWW factor = 3/32 and 1/128.

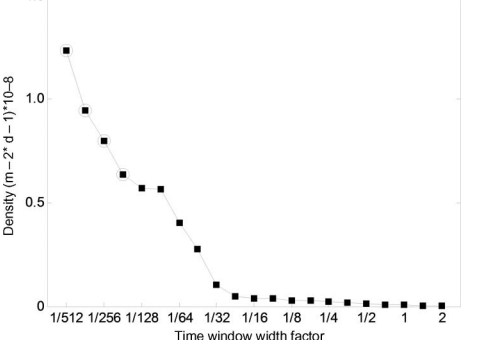

Figure 12. Densities of identified clustered earthquakes versus TWW factor.

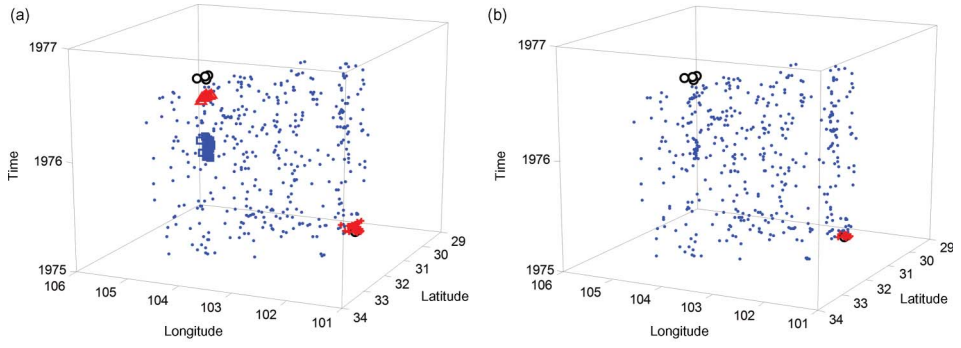

Figure 13. Classification of seismic data using the WNN method: (a) classification generated when TWW factor = 3/32 (strong earthquakes are symbolized by circles; background earthquakes are symbolized by dots; Cluster 1 is symbolized by crosses; Cluster 2 is symbolized by triangles; Cluster 3 is symbolized by squares); (b) classification generated when TWW factor = 1/128 (strong earthquakes are symbolized by circles; background earthquakes are symbolized by dots; Cluster 1 is symbolized by crosses).

Three clusters were identified for the first value (Figure 13a). Are the clustered earthquakes aftershocks or foreshocks? The records of strong earthquakes can help to explain the result. According to Zhang (1990), Cluster 1 (with a higher density) occurred around and after the Daguan earthquake ($M = 7.1$, occurred at 28°06′ N, 104°00′ E on 11 May 1974) and can be treated as the aftershocks of the Daguan earthquake. In addition to the Daguan earthquake, four strong earthquakes (i.e. the Songpan earthquakes; $M = 7.2$, occurred around 32°40′ N, 104°00′ E on 16 August 1976) occurred in and after Cluster 2 and Cluster 3 (these two with a lower density) (Zhang 1986). In this regard, Cluster 2 and Cluster 3 can be treated as the foreshocks and indicative of the Songpan earthquakes. Interestingly, the foreshocks are shown as two distinctive clusters, which we cannot observe if applying spatial clustering methods.

The classification generated when TWW factor = 1/128 is displayed in Figure 13b. From the figure, only aftershocks of the Daguan earthquake (Cluster 1 in Figure 13b) were identified, which is similar to the result of the fourth simulated data set. Both classifications validate that when clusters of varied densities are different in predominant dimensions, a large TWW may help to reveal clusters of different densities whereas a small TWW may only help to reveal clustered earthquakes of a higher density

6. Conclusions and future work

The WNN algorithm was applied to the simulated data and the seismic data in this article and testified as an efficient algorithm for discovering clustering patterns in spatio-temporal data. The novelty of the algorithm lies in three aspects. First, it can not only identify spatio-temporal clusters with arbitrary shapes but also provide fuzzy membership values of events belonging to features. Second, only one parameter, that is, k, needs to be adjusted in the WNN algorithm, because TWW can be determined by locating the local maximum values of density of identified feature events. As a result, WNN is a more objective process compared with other spatio-temporal clustering methods. Third, clusters with different densities and various predominant dimensions can be identified at different scales of TWW. The third point also shows that the WNN algorithm bears analogy to the Fourier transform, in which the components of different frequencies can be extracted through filtering under different thresholds. Although the algorithm was only applied to the seismic catalogue in this article,

we believe it may extend to other areas with respect to spatio-temporal data, such as infectious disease cases and crime venues.

In this article, the algorithm assumes that only two ST Poisson processes exist in a data set, that is, clusters and noise, which means that the algorithm can only deal with two processes simultaneously when TWW is fixed, although the variant of WNN may adapt to data containing more than two Poisson processes. When the number of processes significantly increases, the complexity and the stability of the algorithm will be severely challenged. In addition, only homogeneous point processes are allowed in WNN; inhomogeneous clusters, such as Gaussian process, could lead WNN to generate more false positives or even false clusters. As a result, more sophisticated models, which are capable of dealing with more complex spatio-temporal data, deserve further research.

Acknowledgements

This study was funded through support from the National Key Basic Research and Development Program of China (Project Number: 2006CB701305), a grant from National Natural Science Foundation of China (Project Number: 40830529), the Innovation Project of IGSNRR (Project Number: 200905004) and a grant from the State Key Laboratory of Resource and Environment Information System (Project Number: 088RA400SA).

References

Agrawal, R., *et al.*, 1998. Automatic subspace clustering of high dimensional data for data mining applications. *In*: *Proceedings of 1998 ACM-SIGMOD international conference on management of data*, 4–5 June 1998. New York: ACM Press, 94–105.

Ankerst, M., *et al.*, 1999. OPTICS: Ordering points to identify the clustering structure. *In*: *Proceedings of ACM-SIGMOD'99 international conference on management data*, 27–29 May 1999. Philadelphia, PA: ACM Press, 46–60.

Bastin, L., *et al.*, 2007. Spatial aspects of MRSA epidemiology: a case study using stochastic simulation, kernel estimation and SaTScan. *International Journal of Geographical Information Science*, 21, 811–836.

Birant, D. and Kut, A., 2007. ST-DBSCAN: an algorithm for clustering spatial–temporal data. *Data and Knowledge Engineering*, 60, 208–221.

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

Chen, Y., Liu, J., and Ge, H.K., 1999. Pattern characteristics of foreshock sequences. *Pure and Applied Geophysics*, 155, 395–408.

China Seismograph Network Data Management Center 2009. *China Seismograph Network (CSN) Catalog*. Available from: http://www.csndmc.ac.cn [Accessed 5 January 2009].

Cromwell, P., Olson, J.N., and Avary, D.A.W., 1999. Decision strategies of residential burglars. *In*: P. Cromwell, ed. *In their own words: criminals on crime (an anthology*). Los Angeles: Roxbury, 50–56.

Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 28, 181–187.

Ester, M., *et al.*, 1996. *In*: E. Simoudis, J.W. Han and U.M. Fayyad, eds. *Proceedings of the 2nd international conference on knowledge discovery and data mining*, 3–5 August. Portland: AAAI Press, 226–231.

Feng, H. and Huang, D.Y., 1980. *A catalogue of earthquake in western China (1970–1975, $M \geq 1$)*. Beijing: Seismological Press (in Chinese).

Feng, H. and Huang, D.Y., 1989. *Earthquake catalogue in West China (1976–1979, $M \geq 1$)*. Beijing: Seismological Press (in Chinese).

Gangnon, R.E., 2006. Impact of prior choice on local Bayes factors for cluster detection. *Statistics in Medicine*, 25, 883–895.

Gaudart, J., *et al.*, 2008. Space–time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health*, 6, 1–13.

Grubesic, T.H. and Mack, E.A., 2008. Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, 24, 285–306.

Han, J.W., Kamber, M., and Tung, A.K.H., 2001. Spatial clustering methods in data mining, *In*: H.J. Miller and J.W. Han, eds. *Geographic data mining and knowledge discovery*. London: Taylor and Francis, 188–217.

Hinneburg, A. and Keim, D.A., 1998. An efficient approach to clustering in large multimedia databases with noise *In*: *Proceedings of the fourth international conference on knowledge discovery and data mining*, August. New York, 58–65.

Jacquez, G.M., 1996. A *k* nearest neighbour test for space–time interaction. *Statistics in medicine*, 15, 1935–1949.

Johnson, S.D. and Bowers, K.J., 2004. The stability of space–time clusters of burglary. *British Journal of Criminology*, 44, 55–65.

Karypis, G., Han, E.-H., and Kumar, V., 1999. CHAMELEON: a hierarchical clustering algorithm using dynamic modeling. *IEEE Computer*, 32, 68–75.

Kelsall, J. and Diggle, P., 1995. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*, 14, 2335–2342.

Knox, E.G., 1964. The detection of space–time interactions. *Applied Statistics*, 13, 25–30.

Kulldorff, M., 1997. A spatial scan statistic. *Communications in statistics: theory methods*, 26, 1481–1496.

Kulldorff, M. and Hjalmars, U., 1999. The Knox method and other tests for space–time interaction. *Biometrics*, 55, 544–552.

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

Kulldorff, M., *et al.*, 2005. A space–time permutation scan statistic for disease outbreak detection. *Plos Medicine*, 2, 216–224.

Lian, M., *et al.*, 2007. Using geographic information systems and spatial and space–time scan statistics for a population-based risk analysis of the 2002 equine West Nile epidemic in six contiguous regions of Texas. *International Journal of Health Geographics*, 6, doi:10.1186/1476-072X-6-42.

Mostashari, F., *et al.*, 2003. Dead bird clusters as an early warning system for West Nile virus activity. *Emerging Infectious Diseases*, 9, 641–646.

Nagesh, H., Goil, S., and Choudhary, A., 1999. *Mafia: efficient and scalable subspace clustering for very large data sets*. Technical report TR #9906-010, Northwestern University. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.36.8684 [Accessed January 2009].

Pei, T., *et al.*, 2003. Multi-scale expression of spatial activity anomalies of earthquakes and its indicative significance on the space and time attributes of strong earthquakes. *Acta Seismologica Sinica*, 3, 292–303.

Pei, T., *et al.*, 2006. A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. *International Journal of Geographical Information Science*, 20, 153–168.

Pei, T., *et al.*, 2007. Delineation of support domain of feature in the presence of noise. *Computers and Geosciences*, 33, 952–965.

Pei, T., *et al.*, 2009. DECODE: A new method for discovering clusters of different densities in spatial data. *Data Mining and Knowledge Discovery*, 18, 337–369.

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

Ripepe, M., Piccinini, D., and Chiaraluce, L., 2000. Foreshock sequence of September 26th, 1997 Umbria–Marche earthquakes. *Journal of Seismology*, 4, 387–399.

Sabel, C., *et al.*, 2000. Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland. *Social Science and Medicine*, 50, 1121–1137.

Wang, M., Wang, A.P., and Li, A.B., 2006. Mining spatial–temporal clusters from geo-databases. *Lecture Notes in Artificial Intelligence*, 4093, 263–270.

Wang, W., Yang, J., and Muntz, R., 1997. STING: a statistical information grid approach to spatial data mining. *In*: *Proceeding of the 23rd international conference on very large data bases* (VLDB'97), 25–29 August 1997. Athens: 186–195.

Williams, G.W., 1984. Time space clustering of disease, *In*: R.G. Cornell, ed. *Statistical methods for cancer studies*. New York: Marcel Dekker, 167–227.

Wu, K.T., *et al.*, 1990. *Panorama of seismic sequence*. Beijing: Beijing University Press (in Chinese).

Wyss, M. and Toya, Y., 2000. Is background seismicity produced at a stationary Poissonian rate? *Bulletin of the Seismological Society of America*, 90, 1174–1187.

Yan, P. and Clayton, M.K., 2006. A cluster model for space–time disease counts. *Statistics in Medicine*, 25, 867–881.
Zaliapin, I., *et al.*, 2008. Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, 101, 018501.
Zhang, Z.C., 1986. *Earthquake cases in China (1966–1975)*. Beijing: Seismological Press (in Chinese).
Zhang, Z.C., 1990. *Earthquake cases in China (1976–1980, $M \geq 1$)*. Beijing: Seismological Press (in Chinese).

Appendix

The estimation of the threshold ($D_{\Delta T,k}$) for discriminating between features and noise can be divided into two stages. The first is to construct the mixture pdf of $d_{\Delta T,k}$ of two ST Poisson point processes; the second is to estimate the parameters of the mixture of $d_{\Delta T,k}$ using the EM algorithm.

1. Mixture pdf of WKN distance

According to Assumption 1, noise and features can be represented by two overlapped ST Poisson processes with different intensities, say λ_1 and λ_2; the bimodal pdf of $d_{k,\Delta T}$ can be expressed as follows:

$$d_{\Delta T,k} \sim \rho\Gamma^{1/2}(k, \lambda_{\Delta T,1}\pi) + (1-\rho)\Gamma^{1/2}(k, \lambda_{\Delta T,2}\pi) \tag{A1}$$

where ρ is the proportion coefficient, and $\lambda_{\Delta T,1}$ and $\lambda_{\Delta T,2}$ are the ΔT intensities of the two processes. ρ, $\lambda_{\Delta T,1}$ and $\lambda_{\Delta T,2}$ are the three parameters of the mixture pdf.

2. Estimation of parameters using EM algorithm

The EM algorithm for estimating the parameters of the mixture pdf can be divided into two major steps, namely, the expectation step (E-step) and the maximization step (M-step). The E-step aims at estimating the expectation of the fuzzy membership value of each event subjecting to feature (Byers and Raftery 1998).

The *E*-step is

$$E(\hat{\delta}_i^{(t+1)}) = \frac{\hat{\rho}^{(t)} f_{d_{\Delta T,k}}\left(d_{(\Delta T,k),i};\ \hat{\lambda}_1^{(t)}\right)}{\hat{\rho}^{(t)} f_{d_{\Delta T,k}}\left(d_{(\Delta T,k),i};\ \hat{\lambda}_{\Delta T,1}^{(t)}\right) + (1-\hat{\rho}^{(t)}) f_{d_{k,\Delta T}}\left(d_{(\Delta T,k),i};\ \hat{\lambda}_{\Delta T,2}^{(t)}\right)} \tag{A2}$$

while the *M*-step is

$$\hat{\lambda}_{\Delta T,1}^{(t+1)} = \frac{k\sum_{i=1}^{n}\hat{\delta}_i^{(t+1)}}{\pi\sum_{i=1}^{n} d_{(\Delta T,k),i}^2 \hat{\delta}_i^{(t+1)}} \quad \text{and} \quad \hat{\lambda}_{\Delta T,2}^{(t+1)} = \frac{k\sum_{i=1}^{n}(1-\hat{\delta}_i^{(t+1)})}{\pi\sum_{i=1}^{n} d_{(\Delta T,k),i}^2 (1-\hat{\delta}_i^{(t+1)})}$$
$$\text{with} \quad \hat{\rho}^{(t+1)} = \sum_{i=1}^{n}\frac{\hat{\delta}_i^{(t+1)}}{n} \tag{A3}$$

where n is the number of events, t is the iteration times and $d_{(\Delta T,k),i}$ is the WKN distance of event p_i. If we define the process with $\lambda_{\Delta T,1}$ representing the feature, then events with $\hat{\delta}_i^{(t+1)} \geq 0.5$ belong to feature and events with $\hat{\delta}_i^{(t+1)}$ belong to noise. Here, $\hat{\delta}_i^{(t+1)}$ can be treated as the fuzzy membership value of event p_i belonging to feature.

With all parameters of the mixture pdf of the WKN distance estimated, the threshold ($D_{\Delta T,k}$) for discriminating between feature and noise can be computed with the equation below:

$$D_{\Delta T,k} = \sqrt{\frac{\ln(\frac{1-\rho}{\rho}) + k\ \ln\frac{\lambda_{\Delta T,2}}{\lambda_{\Delta T,1}}}{\pi(\lambda_{\Delta T,2} - \lambda_{\Delta T,1})}} \tag{A4}$$