Geoinformatica (2012) 16:625–652
DOI 10.1007/s10707-012-0165-8

Multi-scale decomposition of point process data

Tao Pei · Jianhuan Gao · Ting Ma · Chenghu Zhou

Received: 10 October 2011 / Revised: 24 April 2012
Accepted: 14 June 2012 / Published online: 3 August 2012

Abstract To automatically identify arbitrarily-shaped clusters in point data, a theory of point process decomposition based on *k*th Nearest Neighbour distance is proposed. We assume that a given set of point data is a mixture of homogeneous processes which can be separated according to their densities. Theoretically, the local density of a point is measured by its *k*th nearest distance. The theory is divided into three parts. First, an objective function of the *k*th nearest distance is constructed, where a point data set is modelled as a mixture of probability density functions (pdf) of different homogeneous processes. Second, we use two different methods to separate the mixture into different distinct pdfs, representing different homogeneous processes. One is the reversible jump Markov Chain Monte Carlo strategy, which simultaneously separates the data into distinct components. The other is the stepwise Expectation-Maximization algorithm, which divides the data progressively into distinct components. The clustering result is a binary tree in which each leaf represents a homogeneous process. Third, distinct clusters are generated from each homogeneous point process according to the density connectivity of the points. We use the Windowed Nearest Neighbour Expectation-Maximization (WNNEM) method to extend the theory and identify the spatiotemporal clusters. Our approach to point processes is similar to wavelet transformation in which any function can be seen as the summation of base wavelet functions. In our theory, any point process data set can be viewed as a mixture of a finite number of homogeneous point processes. The wavelet transform can decompose a function into components of different frequencies while our theory can separate point process data into homogeneous

T. Pei · J. Gao · T. Ma · C. Zhou (✉)
State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, 11A, Datun Road Anwai, Beijing 100101, China
e-mail: zhouch@lreis.ac.cn

T. Pei
e-mail: peit@lreis.ac.cn

J. Gao
e-mail: gaojh@lreis.ac.cn

T. Ma
e-mail: mting@lreis.ac.cn

processes of different densities. Two experiments on synthetic data are provided to illustrate the theory. A case study on reservoir-induced earthquakes is also given to evaluate the theory. The results show the theory clearly reveals spatial point patterns of earthquakes in a reservoir area. The spatiotemporal relationship between the main earthquake and the clustered earthquake (namely, foreshocks and aftershocks) was also revealed.

Keywords Density based clustering method · Homogeneous point process · Wavelet transform · MCMC · EM · *k*th nearest distance

1 Introduction

Point process data are referred to as a set of points in a restricted spatiotemporal area following a certain probability distribution. Many geographic phenomena can be seen as point process data, such as the seismicity in a region over time, distribution of some species, cases of disease in a region within a certain period and spatiotemporal locations of crime in a city. However, most geographic phenomena, affected by various factors at difference scales, are composed of components of different spatiotemporal scales. For example, a seismic data set in an area can be seen as the superposition of background earthquakes (which are distributed over a large spatiotemporal scale with low density) and clustered earthquakes (which are distributed over a small spatiotemporal scale with high density). The separation of different components over different scales may help to identify the spatiotemporal patterns and determine their mechanism. Taking the example of crime, the differentiation between background and clustered crime venues may help to identify crime hotspots and predict the high-crime-incidence seasons and areas. However, because the point data may be very complex because of uneven density, arbitrary shape and space-time coupling, the separation of point patterns over different scales becomes a very difficult problem.

Inspired by Wavelet transform theory, we propose a theory of point process decomposition based on *k*th Nearest Neighbour distance. As we know, with the Wavelet transform, any elementary function can be seen as the summation of wavelet base functions. The components at different frequencies can be extracted using the wavelet coefficients. In our theory, similar to wavelet transformation, the complex point patterns can be seen as a mixture of a finite number of homogeneous (or Poisson) processes with different densities. Here the homogeneous processes can be viewed as the "wavelet base functions" of point process data (as explained in Section 3.2). The area covering the point process (namely, the support domain of point process, defined in Section 3.1) is also similar to the support domain of the wavelet function. If we can find a theory to decompose the point data set into distinct homogeneous processes of different densities, we then realize "a multi-scale decomposition of point process data". With this theory, we can identify the clusters in any point process data set by separating the set into homogeneous components. This paper will discuss the theory, which is constructed by summarizing the approaches to point process data clustering we have made in recent years.

The rest of the paper is organized as follows. Section 2 reviews related work on density clustering methods. Several important concepts are defined in Section 3. Section 4 gives the details of the decomposition theory of point processes. Two experiments of synthetic data are given to illustrate the theory in Section 5. A case study on reservoir-induced earthquakes in Heyuan county, Guangdong province, China, is presented in Section 6. Section 7 provides a summary and directions for future research.

2 Related work

The discovery of clusters of multiple densities has become one of the most important problems in spatial data mining [22, 31, 40]. A good method should meet three standards: (1) it is able to identify clusters of multiple densities; (2) it is capable of handling arbitrarily shaped clusters; (3) the procedure for determining clusters should be objective and require less prior knowledge. To construct a density-based clustering method, we need to solve two key problems. The first is the representation of the point density. The second is to differentiate between clusters of different densities. The recent development of density-based clustering methods is focused on the strategies for solving these two issues. According to the models expressing point density, we classify the density-based clustering methods into five types: (1) grid-based, (2) graph-based, (3) window-based, (4) distance-based and (5) model-based. In the following text, we review these approaches.

The grid-based clustering method classifies points by locating the connected grids where the densities are significantly high. Hinneburg and Keim (1998) proposed the cluster method DENCLUE to identify clusters in a hierarchical tree [23]. Sheikholeslami et al. (1998) proposed the "WaveCluster" method, in which the wavelet transform is employed to determine the clusters from the image generated by counting the number of points in each cell of the grid. Although the "WaveCluster" is capable of identifying clusters over different scales, the detected patterns will lose detail as the scales become coarser [49]. To retain the detail of clusters over larger scales, Murtagh and Starck (1998) adopted the redundant wavelet transform. The method may not only reveal the shapes and locations of clusters over different scales without reducing resolution, but may also eliminate the influence of noise [35]. The main advantage of grid-based clustering methods is their detection capability for finding arbitrary shaped clusters and their high efficiency in dealing with complex data sets which are characterized by large amounts of data, high dimensionality and multiple densities [1, 22]. Nevertheless, the grid-based methods suffer from the major drawback that the clustering results are sensitive to the grid partition scheme, i.e. the size of a grid. The choice of size may significantly affect the outcome of the analysis in terms of size, shape and significance of clusters.

The graph-based methods use graph concepts, such as the *k*th nearest neighbour graph, the Delaunay triangulation and the Voronoi polygon, to classify points. Allard and Fraley (1997) constituted a maximum likelihood estimator for a mixture of uniform point processes using the Voronoi tessellation defined by the data themselves [2]. However, the method is limited to two model components and the specific shape. Karypis et al. (1999) established the clustering method CHAMELEON based on the *k*th nearest neighbour graph [27]. Although CHAMELEON is very good at discovering arbitrarily shaped clusters with multiple densities, more parameters need to be estimated with prior knowledge. Estivill-Castro and Lee (2002) proposed a clustering method (AMOEBA) based on the Delaunay triangulation in which edges have a weight corresponding to the geographical distance between the associated end-vertices [16]. Its aim is to extract the edges whose end-points belong to the same cluster. Typically, graph-based algorithms remove uninteresting edges based on a certain criterion function and then compute the connected components with remaining edges to reveal the clusters [33]. The main weakness of graph-based clustering is its vulnerability to multiple bridges (a bridge is an edge whose deletion increases the number of connected components) [11]. Estivill-Castro and Lee (2002) constructed AUTOCLUST to overcome this following the analysis of Delaunay Diagrams [17]. However, the above methods cannot detect clusters with uneven internal densities. Although Deng et al. (2011) proposed an adaptive spatial clustering algorithm based on Delaunay triangulation (ASCDT)

[11] to identify uneven density clusters, their method needs to compute a series of parameters, which eventually makes the cluster process more subjective.

The window-based method is referred to as the spatial scanning method, which detects regions containing significant clustering points by moving the window [defined as a circle or cylinder (for identifying spatiotemporal clusters)] over the research area [28, 29]. Although the spatial scan statistics have been widely used in disease surveillance, the embedded defect in the method is that the detected features may be significantly influenced by the shape of the window [21, 30, 54]. The choice of an inappropriate window shape may split one feature into many small ones or merge different features into a larger one. In the spatial scan method, the clusters with different densities can be identified by locating significant clustering areas (in all probability clusters or secondary clusters) on the basis of a statistical assumption (such as Poisson or Burnulli assumptions) [36, 37]. Since the spatial scan method can only identify clusters with a predefined shape (usually predefined circles, ellipses or rectangles), more work has been done to extend the method to the detection of irregularly shaped clusters. Duczmal et al. (2007) and Wan et al. (2012) constructed methods for determining arbitrarily shaped clusters of different densities using the Genetic Algorithm (GA) and the Ant Colony Optimization strategy (ACO) respectively [13, 52].

The distance-based clustering method uses the *k*th nearest distance of a point (the distance between the point and its *k*th nearest neighbour) to represent the local density of the point. The representative method is known as the DBSCAN [15]. In DBSCAN, the user estimates the threshold for separating the clusters from noise by locating the "knee" of the *k*-dist graph. Although DBSCAN initially uses the nearest distance to reveal the clusters, it also leaves two unsolved problems: how to identify the clusters of multiple densities and how to estimate the thresholds for clustering points [40]. To adapt the method to multiple densities, some modifications of DBSCAN were developed [32, 34, 38, 46]. Although most of the algorithms mentioned above can deal with data containing clusters with different densities and noise, the estimation of parameters (*MinPts* and *Eps*) is still a subjective process that depends on prior knowledge. To make the estimation of the threshold more precise and objective, Ankerst et al. (1999) proposed the OPTICS algorithm (a variant of DBSCAN), where a reachability plot is constructed to visualize the point connectivity with a one-dimension plot [3]. Daszykowski et al. (2001) proposed a DBSCAN-based modification, which removes the low density points by separating a prescribed percentage of data points, i.e. the tail of the frequency curve [10]. Pei et al. (2006) proposed a nearest-neighbour cluster method, in which the threshold of density (equivalent to *Eps* in DBSCAN) is computed via the Expectation-Maximization (EM) algorithm and the optimum value of *k* (equivalent to *MinPts* in DBSCAN) is decided by the lifetime of the individual *k* [41]. As a result, the clustered points and noise were separated according to the threshold of density and the optimum value of *k*. Although the method can estimate the parameters in an automatic way, it is limited to two-process data sets.

Since most methods are unable to capture the structure of point data fully, Banfield and Raftery (1993) proposed a model-based clustering method to construct the frame for identifying clusters of points [4]. In their method, the clustered events are presumed to be mixtures of Gaussian density and Poisson noise, and any feature can be approximated by the mixture of Gaussian functions. The Bayesian Information Criterion (BIC) is then employed to estimate the number of components. More related work can be found in [19, 20, 25]. Although the Gaussian function has been treated as the base function to accommodate the different components originating from noise and features, it still lacks the ability to describe an arbitrarily shaped cluster, such as tree-like or circle-like clusters. To overcome this limitation, kernel functions were used to reveal statistically significant clustering [12, 51].

Nevertheless, the kernel clustering methods show two limitations: (1) the classification result heavily relies upon the parameter of bandwidth and a different bandwidth will produce a significantly different point pattern; (2) the result of kernel clustering is a smooth surface with many peaks, and there is no exact criterion to determine which cluster relates to which peak [18].

Although all the above methods are capable of separating a point set into different groups according to their differences in density, none of them can meet the three standards (defined in the first paragraph in Section 2) simultaneously. Specifically, the grid-based method is sensitive to the grid size. The graph-based, window-based and most distance-based methods fail to determine the number of components in an objective and automatic way. The model-based method is built on a predefined model (usually Gaussian), so it is unable to discover arbitrarily shaped clusters. In all, there is no theoretical framework that can effectively and objectively determine clusters of multiple densities and arbitrarily shapes in any complex point data set. In this paper, we try to construct such a theory based on our work over the last few years. The theory consists of three parts. The first determines the heterogeneity of a point data set. The second identifies the homogeneous processes in the data set. The third uncovers clusters by grouping density connected points.

3 Several important notes

The essence of the theory is to decompose point process data into different homogeneous processes of different densities. Before we introduce this theory, 8 basic concepts relating to point density are defined for constructing the model. In addition, two fundamental assumptions are presented to lay the foundation of the theory. Readers are referred to [9] for definition 1, [15] for definitions 2 to 6, and [43] for definitions 7 and 8 for more details.

3.1 Some concepts

Definition 1 Support domain: The support domain of a point process is an area that constrains the point process. The probability of a point being distributed at any location in the support domain is defined as the intensity. If the intensity is a constant over the support domain of a point process, then the point process is a homogeneous process.

Definition 2 The *k*th nearest distance of point A ($d_k(A)$): the *k*th nearest distance of point A is defined as the distance between point A and its *k*th nearest neighbour. In this paper, k is also referred to as the distance order.

Definition 3 *Eps*-neighbourhood ($N_{Eps}(O)$): The $N_{Eps(O)}$, indicating the *Eps*-neighbourhood of Point O, is defined by $N_{Eps}(O)=\{O \in D \mid \text{dist}(O,P) \leq Eps\}$. The area in the circle is the *Eps*-neighbourhood of Point O.

Definition 4 Spatiotemporal neighbourhood of point $A(x_i, t_i)$ ($N_{\Delta S,\Delta T}(A)$): a spatiotemporal cylinder centred at A with a geographic radius of ΔS and a temporal window width of ΔT (see Fig. 1).

Definition 5 The windowed *k*th nearest neighbour of a point $A(x_i,t_i)$ ($P_k(x_{i+k},t_{i+k})$): the *k*th geographic nearest neighbour within the spatiotemporal neighbourhood of $A(x_i,t_i)$ ($N_{\infty,\Delta T}(A)$). In Fig. 1, B_5 is the windowed 5th nearest neighbour of Point A.

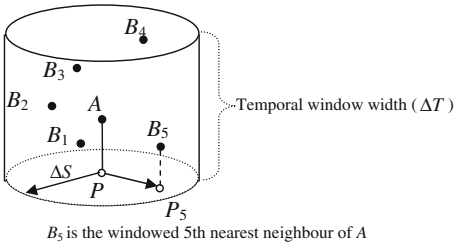


Fig. 1 Spatiotemporal neighbourhood of Point *A*

Definition 6 The windowed *k*th nearest distance of a point A(x_i,t_i) ($d_{k,\Delta T}(A)$): the geographic distance between *A* and its *k*th nearest neighbour $P_k(x_{i+k},t_{i+k})$ in its spatiotemporal neighbourhood ($N_{\infty,\Delta T}(A)$), i.e. $d_{k,\Delta T}(A) = \|x_{i+k} - x_i\|$. In Fig. 1, if we let *P* and P_5 be the projections of *A* and B_5 respectively, then PP_5 is the windowed 5th nearest distance of point *A*.

Definition 7 Spatial density-connected: Point *p* is density-connected to Point *q* regarding *Eps* and *MinPts* if there is a chain of point $p_1, p_2, \ldots, p_n$, $p_1=q$, $p_n=p$ such that $p_{i-1} \in N_{Eps}(p_i)$ (i=2, 3, …, n) and $N_{Eps}(p_i)$ (i=2, 3, …, n-1) must contain at least *Minpts* points. A cluster is defined as a point set *M* that any point $p_i \in M$ is density-connected to point $p_j \in M$ ($p_i \neq p_j$). Thus, noise is the set of points that does not belong to any given cluster. Note that a cluster *M* wrt. *Eps* and *MinPts* contains at least *MinPts* points.

Definition 8 Spatiotemporal density-connected: Point *p* is considered to be spatiotemporal density-connected to Point *q* regarding ΔS,ΔT and *MinPts* if there is a collection of point p_1, p_2,…, p_n (with $p_1=q$ and $p_n=p$) so that $p_{i-1} \in N_{\Delta S,\Delta T}(p_i)$ (i=2, 3, …, n) and $N_{\Delta S,\Delta T}(p_i)$ (i=2, 3, …, n-1) must contain at least *MinPts* points.

3.2 Assumptions relating to point process data

Assumption 1 Because the intensity of an inhomogeneous process can be expressed as a continuous function $\lambda(x)$ [24], the intensity of a local area in the function can be approximated as a constant. As a result, the support domain of all the point process data can be seen as the union of local areas, in each of which the intensity is seen as a constant.

With this assumption, a point process data set can be viewed as approximately the union of a finite number of homogeneous point processes whose support domains are mutually exclusive:

$$S = S_{\lambda_1} \cup S_{\lambda_2} \cup S_{\lambda_3} \ldots \cup S_{\lambda_m} = \bigcup_{i=1}^{m} S_{\lambda_i} \ (m < \infty), \tag{1}$$

where *S* is the whole point set, S_{λ_i} is a homogeneous point process with intensity being λ_i.

Assumption 2 The intensity of an inhomogeneous spatiotemporal (ST) process is a continuous function $\lambda(x,t)$. The intensity of local space in the function can be approximated as a constant. As a result, the support domain of the whole point process data can be seen as the union of local spatiotemporal sets, in each of which the intensity is seen as a constant.

4 The theory of multi-scale decomposition of point process data

There are three subtasks for accomplishing the point process data decomposition. Figure 2 shows the framework of the theory. In the first step, we need to determine if the point process data contain clusters. The second step is to separate the data into homogeneous point processes. There are two different strategies for decomposing the point process data. One is to separate the data into different homogeneous processes based on a mixture density object function of the *k*th nearest distance. If we know the number of point processes in the data, we can use the EM algorithm [41] or Markov Chain Monte Carlo (MCMC) method [55]. If we do not know the number of point processes, we can employ the reversible jump MCMC strategy [40]. The other is to separate the point processes stepwise, that is, to separate the point process data into two parts each time until each part is determined to be a homogeneous process [42]. The third step is to identify clusters in each homogenous point process according to the density connectivity between points. We now describe each step in detail.

4.1 Determining the heterogeneity of point process data

Determining the heterogeneity of point process data once was recognized as a fundamental problem in ecology. Dozens of indices have been developed. The indices are broadly divided into two types. The first is the summary index which uses the property of a point process over a specific scale to detect the point pattern as a whole [14, 26, 38, 44, 48]. The second is the scale indices which are capable of revealing the point pattern over various scales, such as the *L*-function and the *K*-function and the ratio of the variance of the *k*th nearest distance [5, 45]. Since the scale indices may provide more data information relating to the scale, they

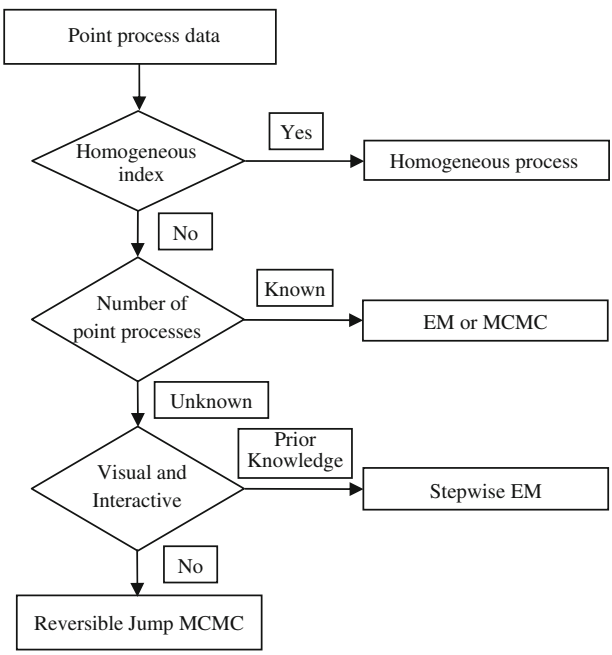


Fig. 2 Framework of point process data decomposition

have been increasingly used in different disciplines [9]. Nevertheless, previous scale indices show two limitations, which make the computation of scale indices a compute-intensive and subjective process. First, the thresholds for determining the heterogeneity of data are derived from the statistics of simulations. Second, some parameters, for instance the intensity of the process, should be estimated beforehand in computing the indices [9]. To enhance the detection power, we developed a non-parameter index to determine the heterogeneity of a point process data set based on Assumption 1 [39]. The index is defined as:

$$\frac{Var^*_{k+1}(x)}{Var^*_k(x)}/R_k, \tag{2}$$

where $Var^*_{k+1}(x)$ ($Var^*_k(x)$) is the variance of the (k+1)th (kth) nearest distance for the point process data, R_k is a constant (the ratio between the expectation of the variance of the (k+1)th nearest distance and that of the kth nearest distance for a homogeneous point process). The data set contains clusters if $\frac{Var^*_{k+1}(x)}{Var^*_k(x)}/R_k > 1 \quad (1 \le k \le K)$, where K is smaller than the number of points in the clusters. Usually, we can set K to a value between 5 and 10. For more details refer to [39]. That is, if $\frac{Var^*_{k+1}(x)}{Var^*_k(x)}/R_k > 1$ as k is smaller than, say 5, then the data must contain clusters; otherwise, they are homogeneous.

4.2 Likelihood function of the *k*th nearest distance

Before decomposing a point process data set into different homogeneous point processes, we should introduce the likelihood function of the kth nearest distance of the data. We then derive the probability density function (pdf) of the kth nearest distance for the 2D Poisson process. Given a point p_i in the 2D Poisson point set Y, the spatial probability distribution of its kth nearest distance D_k (the distance between p_i and its kth nearest neighbour) can be acquired by seeking the pdf of all 0, 1, 2, …, k-1 points within the circle of $A(p_i, x)$, in which p_i is the centre and x is the radius [6].

$$P(D_k \ge x) = \sum_{m=0}^{k-1} \frac{e^{-\lambda\pi x^2}(\lambda\pi x^2)^m}{m!} = 1 - F_{D_k}(x) = 1 - P(D_k < x). \tag{3}$$

When a point data set contains m point processes S_{λ_i} (i=1,2,…m), the likelihood function of the parameters of the mixture can be written as:

$$D_k \sim \sum_{i=1}^{m} p_i f(x; k, \lambda_i) = \sum_{i=1}^{m} p_i \frac{e^{-\lambda_i\pi x^2} 2(\lambda_i\pi)^k x^{2k-1}}{(k-1)!}, \tag{4}$$

where $p_i \ge 0$ is the proportion of the ith component with $\sum p_i = 1$, m is the distance order and λ_i is the intensity of S_{λ_i} [40].

4.3 Strategies for estimating the parameters of the spatial point process

(1) The EM algorithm

The EM algorithm can be employed if the number of homogeneous processes in the data is known. The EM algorithm is an iterative algorithm for estimating the parameters of the mixture pdf. It contains two steps, the first is the Expectation step (E-step):

$$E\left(\widehat{\delta}_{i,j}^{(t+1)}\right) = \frac{\widehat{p}_j^{(t)} f_{D_k}\left(d_i; k, \widehat{\lambda}_j^{(t)}\right)}{\sum_{j=1}^{m} \widehat{p}_j^{(t)} f_{D_k}\left(d_i; k, \widehat{\lambda}_j^{(t)}\right)} \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, m). \tag{5}$$

The second is the Maximization step (the M-step):

$$\widehat{\lambda}_j^{(t+1)} = \frac{k \sum_{i=1}^{n} \widehat{\delta}_{i,j}^{(t+1)}}{\pi \sum_{i=1}^{n} d_i^2 \widehat{\delta}_{i,j}^{(t+1)}} \; with \, p_j{}^{(t+1)} = \sum_{i=1}^{n} \widehat{\delta}_{i,j}^{(t+1)} / n \quad (i = 1, 2, ..., n; j = 1, 2, ..., m), \tag{6}$$

where p_i and λ_i share the same meaning as those in Eq. 4, n is the number of points, t is the iteration time, d_i is the kth nearest distance of point i, $\widehat{\delta}_{i,j}^{(t+1)}$ can be seen as the probability of point i belonging to point process j. The M-step is computed with the result of the E-step as its input (i.e. $\widehat{\delta}_{i,j}^{(t+1)}$), and vice versa. The parameters will be determined iteratively until both steps have converged [6, 41].

(2) Reversible jump MCMC algorithm

The reversible jump MCMC strategy is used to estimate the parameters (m, $\lambda_{1:k}, p_{1:k}$) if the number of the point processes in the data is unknown. In this strategy, λ_i is simulated with a log-normal random walk while p_i is simulated with an additive normal random walk on the logit scale. Among the parameters, m is the most difficult to estimate. Here we use a birth-and-death step, where a random choice between birth and death is made through a Markov Chain simulation process. For more details refer to [40]. The advantage of the reversible jump MCMC is that the distinct homogeneous processes can be separated simultaneously comparing to the Stepwise EM algorithm (described in the next section). The disadvantage is that the algorithm is complex and time-consuming.

(3) Stepwise EM algorithm

The main idea of the stepwise EM algorithm is to treat the point process data as two components and separate the data into noise and clusters in each step. In detail, we first separate a point process data set into noise and clusters. Then we treat the noise or clusters as a new data set and reapply the same procedure until the final remnants are determined as homogeneous processes or the number of remaining points is smaller than a certain threshold [42]. Here we can use the index in formula (2) to determine the homogeneity of the intermediate results generated at each step. There are two advantages of stepwise EM. First, we do not need to know the number of components in the data set. Second, unlike reversible jump MCMC, the algorithm is easy to realize and has relatively low computational complexity. The disadvantage is that users must determine if the result is homogeneous at each step. If the temporary data is inhomogeneous, the procedure will continue.

4.4 The choice of k

In the above algorithms, the only parameter is k. The choice of k may significantly affect the result. As k is small, the components (i.e. the homogeneous point process) in the histogram of the kth nearest distance will not be clearly parted, which might generate false positives. As k is too large, on one hand, the algorithm will miss the patterns which contain fewer points than k; on the other hand, because the kth nearest distances of points (either in noise or in clusters) near the border between clusters and noise may be smaller (for noise) or larger (for clusters) than expected, the mixture histogram of the kth nearest distance will deviate from

the theoretical one, which may cause the clusters (noise) to be overestimated (underestimated). As a result, the value of k should be neither too large nor too small. We suggest a value of k between 5 and 20.

4.5 Mechanism for connecting points into clusters

Before connecting points into clusters, we estimate the thresholds of the kth nearest distance for differentiating between clusters. The formula for estimating the thresholds is derived by determining the x coordinate of the intersection of the pdfs between S_{λ_i} and S_{λ_i+1} :

$$Eps_i^* = \sqrt{\frac{\ln \frac{p_i}{p_{i+1}} + m \ln \frac{\lambda_i}{\lambda_{i+1}}}{\pi(\lambda_i - \lambda_{i+1})}} \quad (i = 1, 2, \ldots m-1), \tag{7}$$

where p_i shares the same meaning as in Eq. 4, Eps_i^* is the threshold between the ith point process and the $(i+1)$th, that is, if the kth nearest distance of a point is between Eps_i^* and Eps_{i+1}^* , then the point belongs to the process S_{λ_i} . With Eq. 7, we will obtain m-1 thresholds, where m is the number of homogenous point processes in the data set. The mechanism of extending clusters of different densities is same as that in DBSCAN [15, 47]. In detail, the construction of a cluster is a recursive process which connects the spatial density-connected points with respect to $Eps_i^*\ (i = 1, 2, \ldots m-1)$ and k (Definition 7). All clusters are generated until no spatial density-connected point is left in the data. Each pair of parameters (Eps_i^*, k) will generate a set of clusters with the same density.

4.6 Windowed Nearest Neighbour Expectation Maximization (WNNEM) algorithm

When the data contain temporal information, we can use the WNNEM method to determine the clusters. With the windowed kth nearest distance ($D_{\Delta T,k}$), we can identify the spatiotemporal clusters. The likelihood of the parameters sharing the same formation as Eq. 4 is:

$$D_{\Delta T,k} \sim \sum_{i=1}^{m} p_i f\left(x; k, \lambda_{\Delta T,i}\right) = \sum_{i=1}^{m} p_i \frac{e^{-\lambda_{\Delta T,i}\pi x^2} 2\left(\lambda_{\Delta T,i}\pi\right)^k x^{2k-1}}{(k-1)!}, \tag{8}$$

where $p_i \geq 0$ is the proportion of the ith component with $\sum p_i = 1$, k is the distance order and $\lambda_{\Delta T,i}$ is the intensity of the ith spatiotemporal homogeneous point process [43].

All methods for estimating the parameters can apply to the spatiotemporal scenario. The only point to be noted is the choice of ΔT. Different values of ΔT will generate different clustering results. For details, readers may refer to [43]. Here we give only a brief introduction. We first calculate the densities of clusters at various values of ΔT. The values, at which the average density of clustered points reaches its local maxima, can then be used as the temporal window width. The reason for this is that if less noise is classified into the clusters the density of clustered points will reach its maximum. As a result, a precise result will be generated when using the threshold obtained at the maximum. When there are multiple local maxima of the densities of clustered points, multiple values of ΔT will be obtained and used to mine the clusters.

When ΔT is no smaller than the whole temporal width of all points, the spatiotemporal clustering becomes spatial clustering. When ΔT is large, the method will identify clusters extended in the T-direction (i.e. time). The reason is that in a cluster extended in the T-direction the windowed kth nearest distances of points are smaller on average than those of

points in noise or clusters extended in the X-Y plane within such a larger ΔT. In other words, it is easier for the points in the cluster extended in the T-direction to find the k nearest neighbours than those in the cluster extended in the X-Y plane. When ΔT decreases, the method will identify clusters extended in the X-Y plane. This is because in a cluster extended in the X-Y plane the windowed kth nearest distance of a point is smaller than that of points in the noise or clusters extended in the T-direction within such a small ΔT. As a result, the cluster extended in the X-Y plane will be correctly identified as a cluster since ΔT is small.

5 The experiment of simulated data

We use two synthetic data sets to validate our theory. In the first experiment, the data are composed of several distinct homogeneous point processes. In the second experiment, the data are distributed as a heterogeneous process, where the intensity of the point process is not a constant but a continuous function, in other words, the density of points are smoothly changed over its support domain.

5.1 Experiment 1

The first data set, shown in Fig. 3, is distributed in 3 dimensions (specifically, Axes X and Y indicate the spatial coordinates while T indicates the temporal coordinate) and composed of noise and four clusters of different densities. In detail, the clustered points contain one circular-disk cluster (Cluster 1), one cylinder cluster (Cluster 2), one L-shaped cluster (Cluster 3) and one thin plate cluster (Cluster 4). Note that Cluster 1 is embedded in Cluster 2. The spatiotemporal densities of clusters, in descending order, are Cluster 1, Cluster 2, Cluster 3 and Cluster 4. When only considering spatial coordinates, Cluster 1 and Cluster 2 have the highest density and then Cluster 3, followed by Cluster 4. We then use the reversible jump MCMC and stepwise EM to decompose the data in 2D (namely, clustering the points in the X-Y plane) and WNNEM in 3D.

The ratio of the variance of the kth nearest distance was used first to determine the heterogeneity of the data. $\frac{Var^*_{k+1}(x)}{Var^*_k(x)}/R_k$ is larger than 1 as $k<10$, indicating the point data are inhomogeneous. By setting k to 15, we generated the result using the reversible jump MCMC (Fig. 4). Figure 4a demonstrates that the mixture pdf of the kth nearest distance consists of three components, each of which represents a homogeneous point process. The densities of the point processes decrease from left to right (because the kth nearest distance increases from left to right). Three clusters and noise, subjected to three different point processes, are generated (Fig. 4b). Specifically, Cluster 1, Cluster 2 and Cluster 3 are successfully identified. Note that Cluster 1 and Cluster 3 are in the same point process. Cluster 4 is not identified because the points in it are not concentrated in the X-Y plane.

The result generated by stepwise-EM is displayed in Fig. 5. Three clusters and noise are identified within two steps. In detail, noise is separated in the first step. In the second step, Cluster 1 is separated from Cluster 2 while Cluster 3 is identified by its homogeneity. The algorithm stops because Cluster 1 and Cluster 2 are both determined as homogeneous. Differing from the result generated by MCMC, Cluster 2 (some points are classified into noise) is underestimated by Stepwise EM.

We used WNNEM to identify spatiotemporal clusters from the data. In Fig. 6 the densities of the clustered points are plotted over the temporal window factor (ΔT can be obtained by multiplying the factor by the whole temporal width of the data), from which we

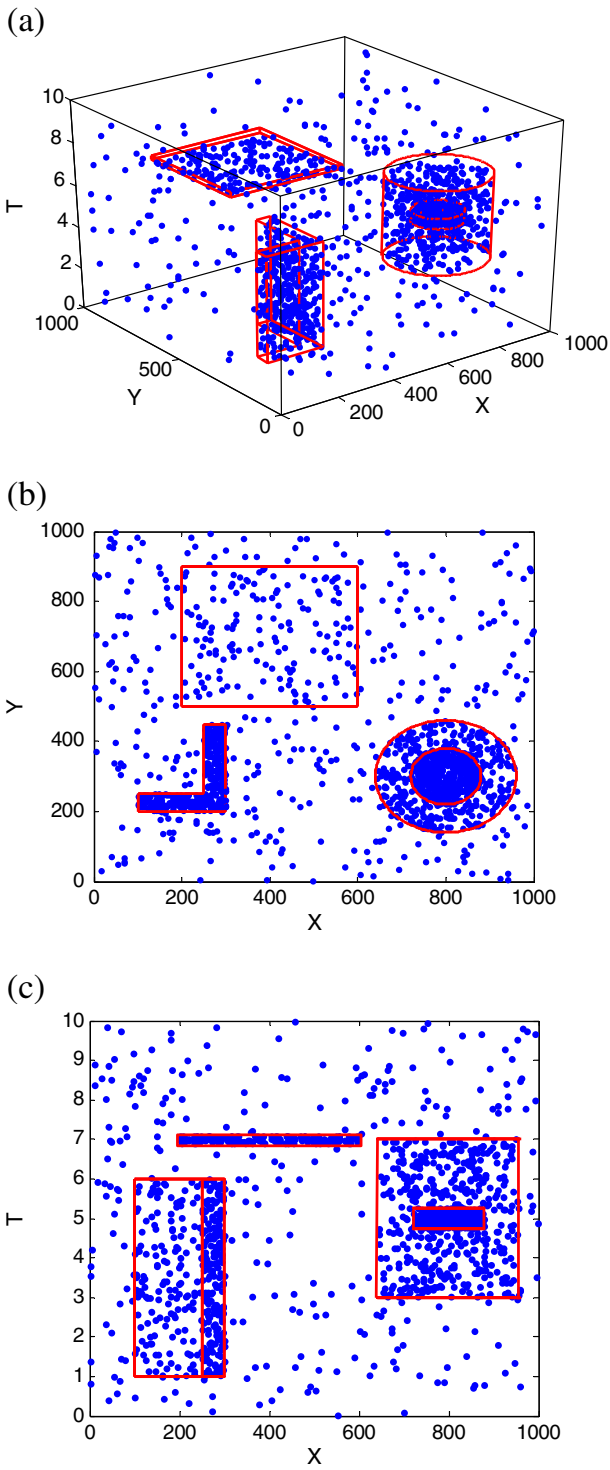


Fig. 3 Simulated data of Experiment 1. **a** 3D view; **b** X-Y plane; **c** X-T plane

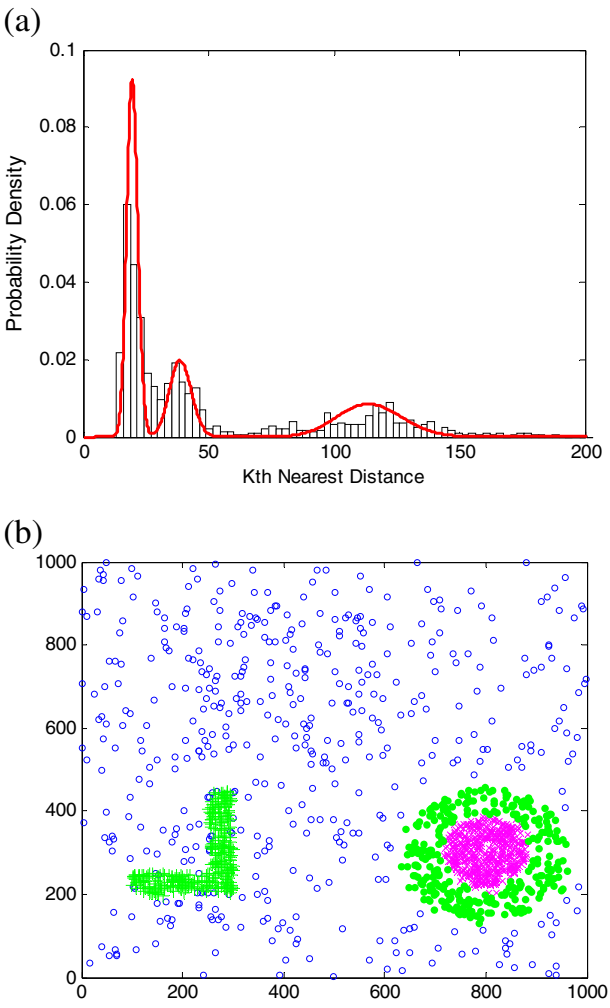


Fig. 4 Clustering result of reversible jump MCMC. **a** Histogram of the *k*th nearest distance and its fitted curve; **b** clusters generated by the reversible jump MCMC (Cluster 1 is indicated by crosses; Cluster 2 by dots; Cluster 3 by plus signs and noise by circles)

found three local maxima (i.e. 1/512, 3/512 and 3/16). We then used different ΔTs, generated at different maxima of temporal window factors, to determine the clusters in the data (Fig. 7). As ΔT=1/512, Cluster 1 is identified while other points are classified as noise (Fig. 7a). As ΔT=3/512, Cluster 1 and the Cluster 4 are both discovered (Fig. 7b). As ΔT=3/16, Cluster 1 and Cluster 2 are combined (i.e. Cluster 2′, indicated by dots), while Cluster 3 is also identified. Note that Cluster 4 is classified as noise again (Fig. 7c).

The spatial clustering results demonstrate that the reversible jump MCMC and the stepwise EM is able to identify the clusters. Nevertheless, because in each step only two components are assumed to exist in the data in stepwise EM, some points may be misclassified, which leads to underestimation (or overestimation) of clusters, say Cluster 2. On the contrary, the reversible jump MCMC can generate more precise results because it can separate the point processes simultaneously. Because the reversible jump MCMC and the stepwise EM can only deal with spatial points, they may miss some spatiotemporal clusters which do not appear to cluster in the X-Y plane (for example, they fail to uncover Cluster 4).

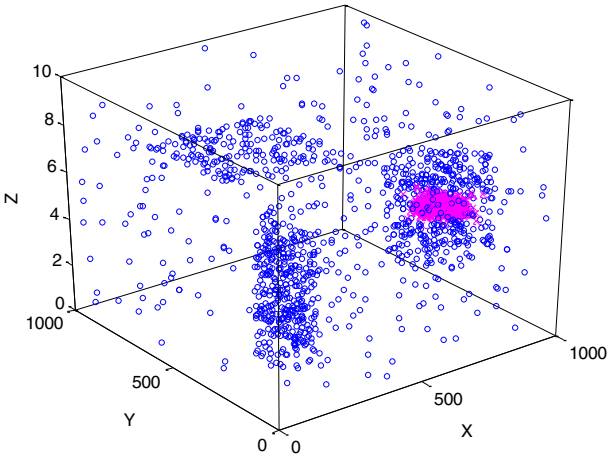


Fig. 5 Clustering results of Stepwise EM. (Cluster 1 is indicated by crosses; Cluster 2 by dots; Cluster 3 by plus signs and noise by circles)

The WNNEM can identify spatiotemporal clusters of different dominant directions and densities by using different values of ΔT, which can be indicated by the local maxima of the spatiotemporal intensity of clustered points.

5.2 Experiment 2

The intensity, used to generate the heterogeneous point process data, was calculated with the function of $\lambda(x) \propto \rho(x-1/2)^2 + 1/2$ [where x is the horizontal location of a point and ρ is the trend constant (Complete Spatial Randomness if ρ=0)] [7]. Here we let ρ=10. Function $\lambda(x)$ makes the points become denser horizontally from the central part to the edges. The heterogeneous data are displayed in Fig. 8a. We only use the reversible jump MCMC to decompose the data. The histogram of the kth nearest distance (here we let k=15) demonstrates that there are two dominant components in the data, namely the clustering and background processes. Figure 8b displays the points that are divided into two clusters (these two clusters belong to the clustering process) and the background. We can see that the clusters (i.e. the process of high intensity) are distributed in the left and right sides of the

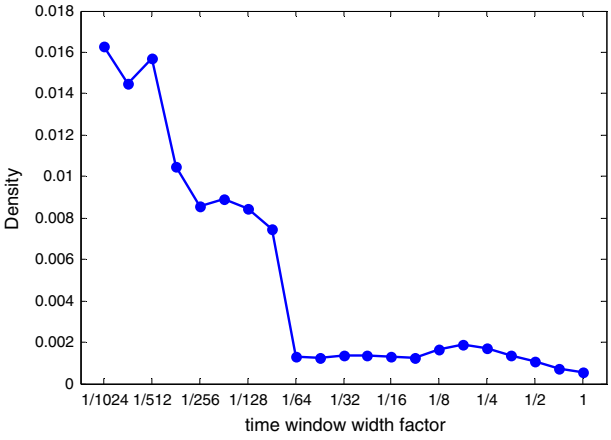


Fig. 6 Density of clustered points over different time window width factors

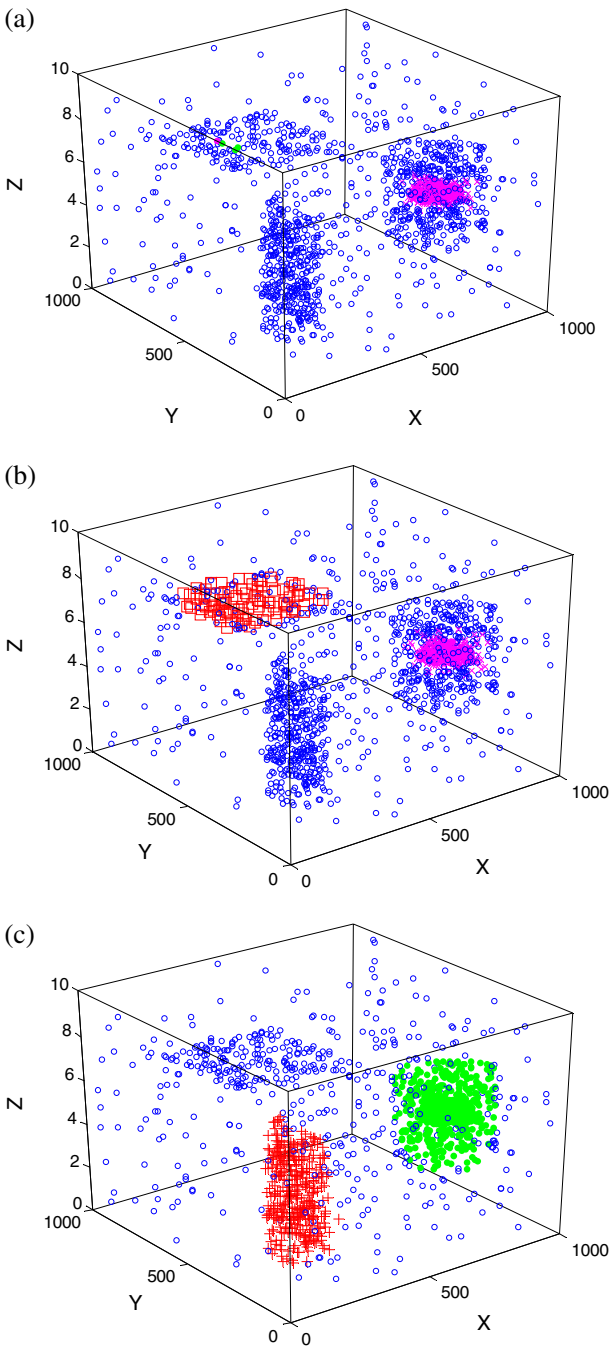


Fig. 7 Results of WNNEM generated for different time window width factors. **a** 1/512 (Cluster 1 is indicated by crosses); **b** 3/512 (Cluster 1 by crosses and Cluster 4 by rectangles); **c** 3/16 (Cluster 3 by plus signs and Cluster 2′ by dots)

region, which conforms the intensity value distribution generated by $\lambda(x)$. Because the density of points in the heterogeneous point process changed smoothly over the support domain, the reversible jump MCMC can only provide an approximate solution, which

contains a number of distinct homogeneous point processes, to discern the point subsets of high density from those of low density. In this experiment, the data are divided into two point process subsets.

6 A case study on reservoir-induced earthquakes

We use a case study of reservoir-induced earthquakes to validate the point process decomposition theory. The research area was in Heyuan County, Guangdong Province, China (Fig. 9), where the Xinfengjiang reservoir is located. The Xinfengjiang Reservoir was impounded in 1959. After impounding, the number of earthquakes in the reservoir area increased dramatically [50]. On 19 March, 1962, a strong earthquake, measured as 6.1 M, occurred in the reservoir and caused serious damage to the dam. Previous studies have shown that: (1) the increase in the number of earthquakes is related to the reservoir impounding, (2) early clustered earthquakes (namely, foreshocks) could indicate the occurrence of the main strong quake, and (3) the occurrence of the main strong quake also induced more aftershocks[50, 53]. In this case study, we used the point process decomposition theory to reveal the above relationships. The seismic records between 7 July 1961 and 8 July 1969 were collected from the China Seismograph Network (CSN) Catalogue [8]. All records were measured on the Richter scale and had magnitudes greater than 2.

6.1 Identification of spatially clustered earthquakes

In this case study, we first used two methods (the reversible jump MCMC and the stepwise EM) to identify the spatial patterns of the seismicity. The results generated by the two methods were then compared to determine their differences. Finally, we used the WNNEM method to determine the spatiotemporal clusters from the data as well as the relationship between the clustered earthquakes and the main strong earthquake.

(1) Results of reversible jump MCMC

We used the ratio of the variance of the kth nearest distance to determine the heterogeneity of the data. The results indicate that the data are not homogeneous (we omit the result for brevity). The histogram of the 20th nearest distance (note in the case study we set k to 20 for all algorithms) of epicentres is shown as multimodal in Fig. 10, which also confirms that the data are inhomogeneous. The result (the fitted curve in Fig. 10) generated by the reversible MCMC indicates that the data are divided into four homogeneous point processes.

The clustering results with the reversible jump MCMC are displayed in Fig. 11. Note that the same point process is indicated by the same colour. The data are further grouped into 6 point clusters and noise. Note that different clusters are indicated by different symbols. The clustered earthquakes demonstrate the nested structure. In detail, Process 1 (in red), distributed over the whole reservoir area, can be treated as the background earthquakes over a large scale. Process 2 (in green), located among the background earthquakes, is composed of two clusters (i.e. Cluster M1 and Cluster M2). Process 3 (in blue), surrounded by Cluster M1, consists of two clusters (Cluster M3 and Cluster M4). Process 4 (in yellow), embedded in Cluster M3, also contains two clusters, that is, Cluster M5 and Cluster M6. The nested structure of the decomposition is shown in Fig. 12.

We then analysed the clustering patterns generated by the reversible jump MCMC. In the decomposition result, there are two clustered areas. One is located in the northwest part of the reservoir area (Area X), including Cluster M2. The other is located in the southeast part

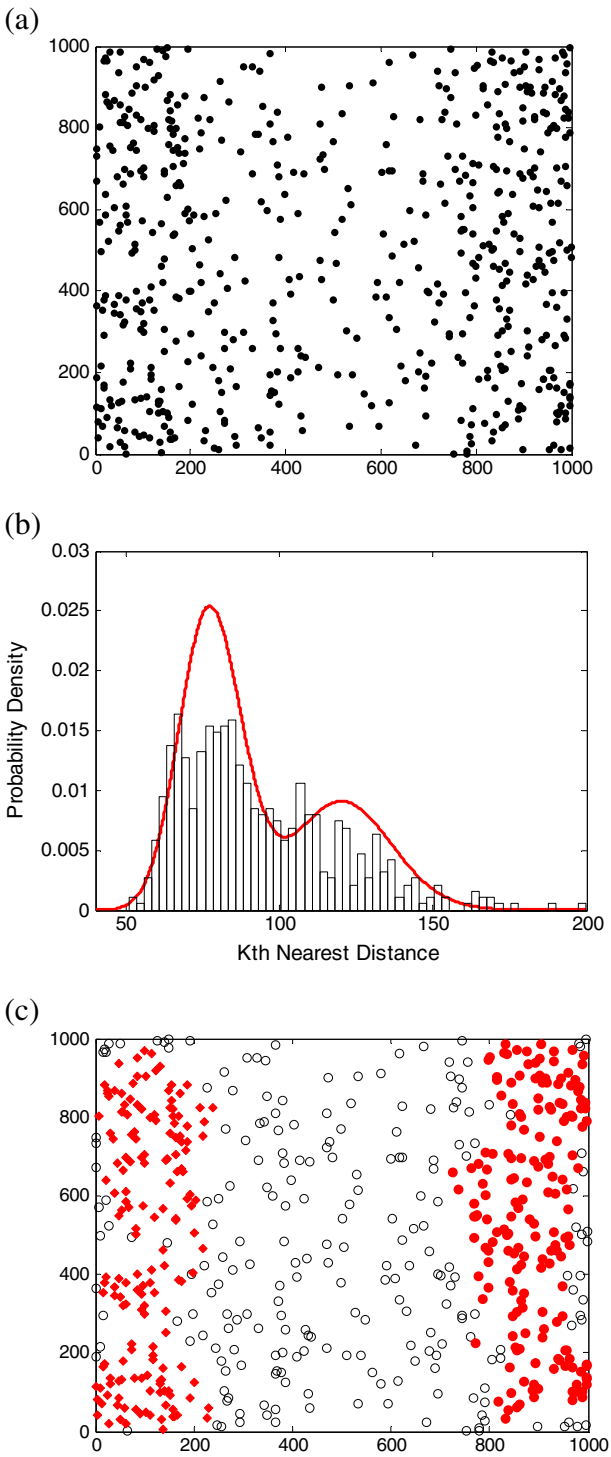


Fig. 8 Heterogeneous point process decomposition by reversible jump MCMC. **a** Original data; **b** The histogram of the *k*th nearest distance and its fitted curve; **c** Clusters and noise. (Cluster 1 is indicated by diamonds, Cluster 2 by dots and noise by circles)

of the research area, including Clusters M1, M3, M4, M5 and M6. The second area contains two clustered subareas over a smaller scale. One is located in the northeast, covering Cluster M4 (Area Y). The other is located in the south (Area Z), covering Clusters M3, M5 and M6. The nested clustering structure conforms to the previous research results on the reservoir induced earthquakes [53]. The seismicity increased after the reservoir was impounded. Before the main quake occurred, as the tectonic stress is in the NW-SE direction, the earthquakes were distributed along the NW-SE direction and concentrated on Areas X and Z [53]. After the main quake (M=6.1) occurred, the tectonic stress was redistributed which caused the generation of new tectonic stress directions [50, 53]. The outcome is that seismicity in Area X weakened, while clustered earthquakes in Areas Y and Z increased. In all, the layout of the clusters could reflect the distribution of tectonic stress and its change over different scales. However, since no temporal attribute was considered in this experiment, we were unable to identify the foreshocks and aftershocks. In addition, it was difficult to interpret the appearance of Clusters M5 and M6.

(2) Stepwise EM

We then applied the stepwise EM algorithm to the same data. The whole separation process is displayed as a binary tree in Fig. 13. Three steps were used to classify the data into four processes, indicated by the same colours, in Fig. 14. From the results, we found that the clustering structure was similar to that generated by the reversible jump MCMC method except for two small differences. The first was that, in the result for the reversible jump of MCMC, there were two medium clusters (M5 and M6) embedded in Cluster M3, while in the results of the stepwise EM there were one large cluster and one small cluster (E5 and E6) embedded in Cluster E3. The second was that although almost all clusters in the stepwise EM results were replicated in the reversible MCMC (they share the same cluster ID in Figs. 11 and 14), they contained different numbers of earthquakes; for instance, Cluster E2 (in the results for the stepwise EM) had more earthquakes than its counterpart (Cluster M2) while Cluster E3 had fewer earthquakes than in Cluster M3.

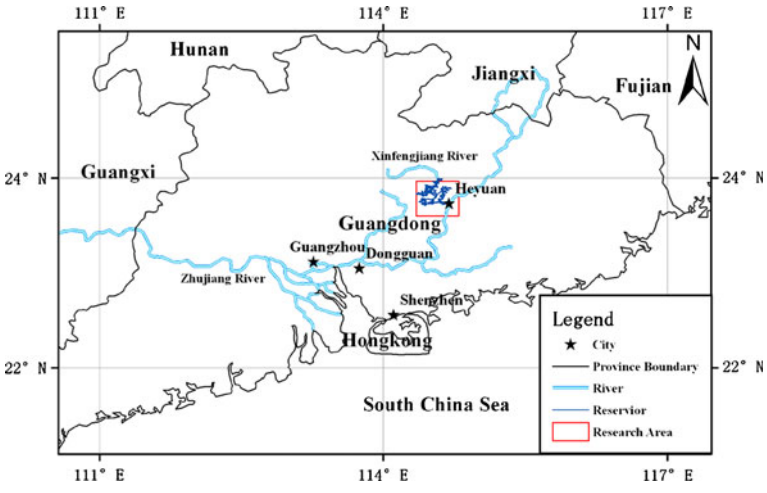


Fig. 9 Location of the research area

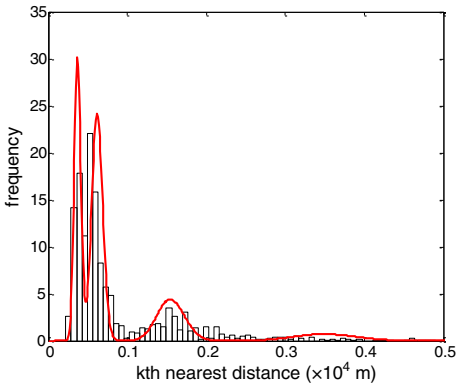


Fig. 10 Histogram of the *k*th nearest distance of the epicentres and its fitted mixture pdf

6.2 Identification of spatiotemporal clustered earthquakes

We then employed the WNNEM method to identify the spatiotemporal clusters. We used the same set of data in the following computation. We first computed the densities of clustered earthquakes over different time window widths and found two maxima (Fig. 15). The first is for one year and the second is for 60 days. We also used the temporal width of all the data (that is, 8 years) to show its spatiotemporal structure (although it is the same as the spatial clustering result). All classification results are displayed in Figs. 16, 17 and 18.

(1) Temporal window = 8 years

The results in the 3D view are displayed in Fig. 16a and their planar results can be seen in Fig. 16b. Clusters L1 and L2 (namely the subset B in the result for the stepwise EM) were generated as the temporal window was set to the largest value. Cluster L1 (i.e. the union of Clusters E3, E5 and E6), indicated by red crosses, extended along the time axis, indicating that the seismicity was intensive in this area during the whole period. Cluster L2 (i.e. Cluster E4), indicated by triangles, was found to appear immediately after the occurrence of the

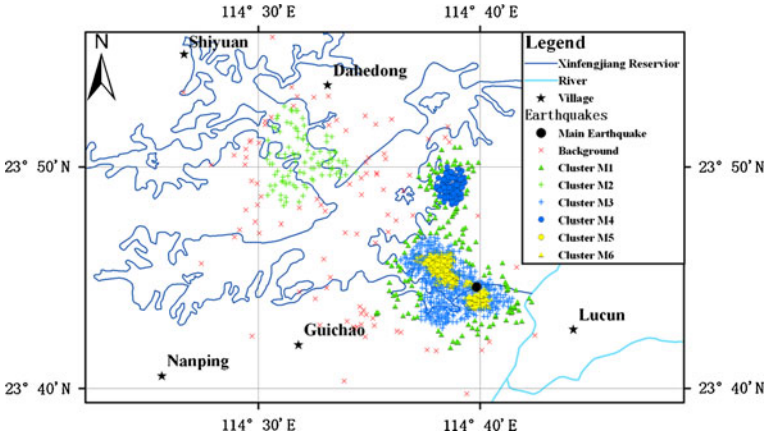


Fig. 11 Decomposition results of reversible jump MCMC

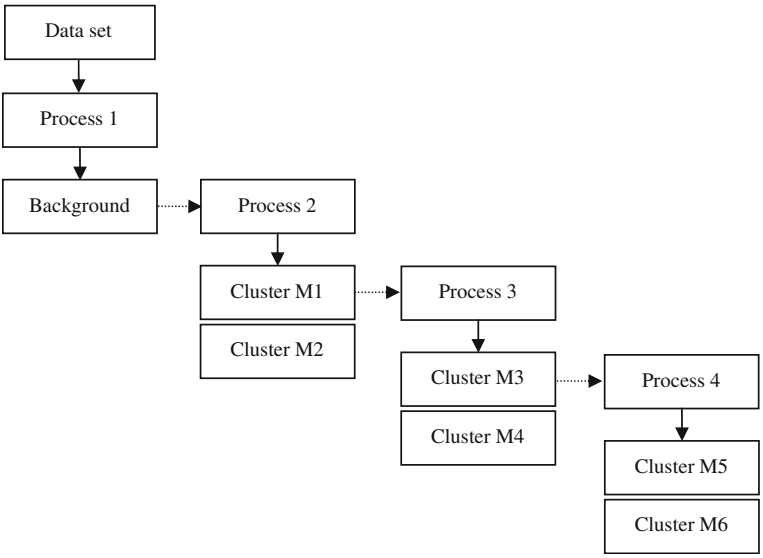


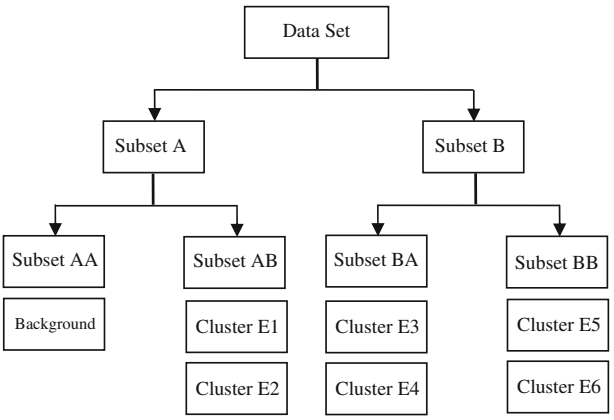
Fig. 12 Nested structure of the decomposition results generated by reversible jump MCMC

main earthquake (Fig. 16a). Here, L2 corresponds to Area Y generated by the reversible jump MCMC while L1 corresponds to Area Z.

(2) Temporal window = 1 year

When the temporal window was decreased to 1 year, clusters I1 and I2 were generated (Fig. 17). Two disparities were found from the comparison with that generated for a temporal window width = 8 year. First, the support domain of Cluster L1 and that of Cluster L2 merged into the support domain of one larger cluster (i.e. Cluster I1) (Fig. 17b). Second, when the temporal window was set to 1 year, Cluster I2, located northwest of Cluster I1, was newly identified. From Fig. 17a, we found that this cluster was located within a short temporal interval.

Fig. 13 Decomposition process with stepwise EM



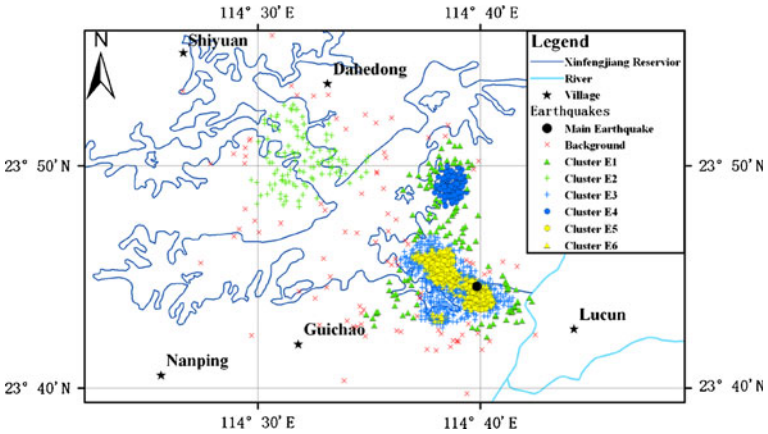


Fig. 14 Decomposition results generated using stepwise EM

(3) Temporal window = 60 days

When the temporal window was further decreased, clusters S1, S2, S3, S4 and S5 were generated. Most were concentrated in narrow temporal intervals (Fig. 18a). Specifically, Clusters S3, S4 and S5 spatially overlapped with Cluster S1 (Fig. 18b). Interestingly, Cluster S3 was located before the occurrence of main earthquake of Xinfengjiang (6.1 M). Clusters S4 and S5 are two thin clusters. Cluster S4 was located above Cluster S1 while Cluster S5 was located above all clusters (Fig. 18a).

From the above analysis, we find that different window widths may generate clusters, which are prolonged in different dimensions. Specifically, the narrow window will generate the clusters extended in the XY plane while the wide window will produce clusters prolonged along the temporal axis. The results generated for a temporal window = 8 years

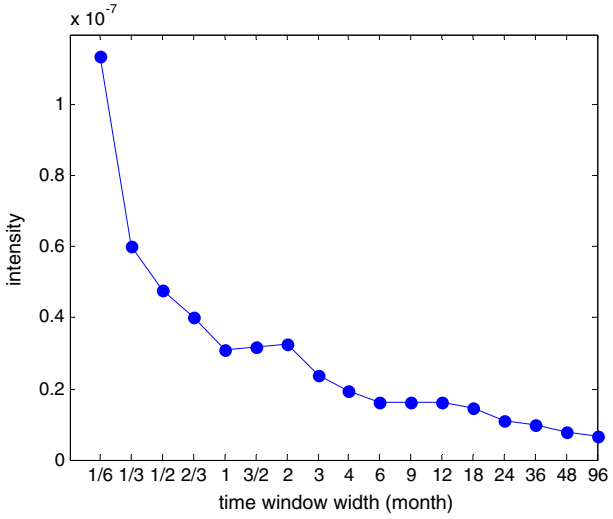


Fig. 15 Intensity of clustered earthquakes for different time window widths

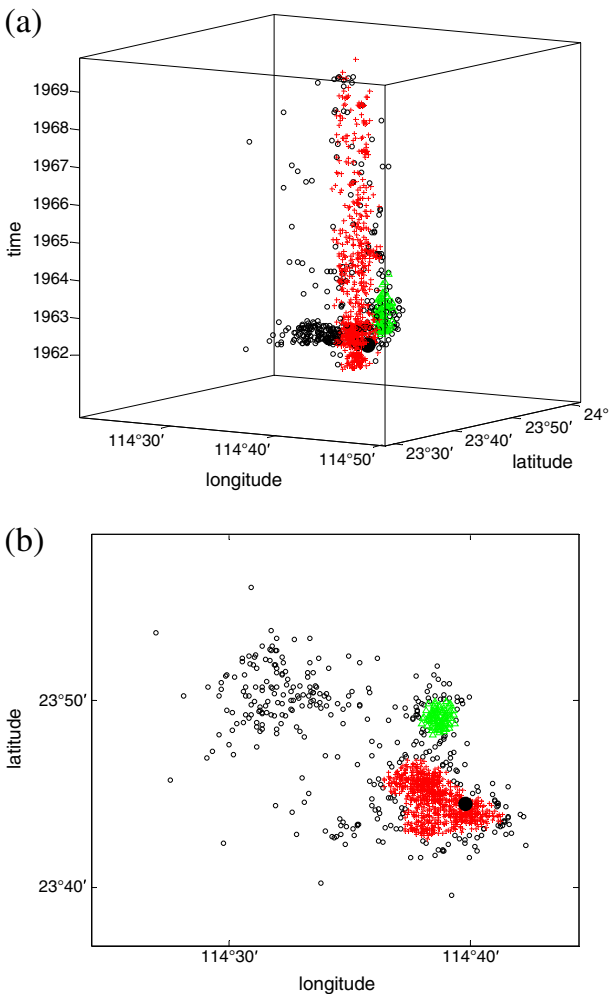


Fig. 16 Clustering results generated for a temporal window = 8 year. (**a**) 3D view; (**b**) 2D view (Cluster L1 is indicated by red crosses; Cluster L2 by green triangles; noise by black circles and the main earthquake by a larger black dot)

only revealed the two spatial clusters (L1 and L2). Although these two clusters were recognized as the most intensive seismicity areas, no temporal structure concerning the clusters was revealed. As the temporal window width decreased, the support domains of Clusters L1 and L2 merged into a larger one, which turned into Cluster I1. Because of the temporal window width = 60 days, more temporally concentrated clusters were identified. Cluster S3 was separated from Cluster L1 and can be regarded as the foreshocks of the main earthquake, which indicated the spatiotemporal location of the main earthquake.

Compared with spatial clustering results, the spatiotemporal clustering results allowed us to identify the temporal order of different clusters. There were five new discoveries with the windowed *k*th nearest distance. First, the foreshocks (i.e. Cluster S3) were identified by the spatiotemporal clustering method for the temporal window width = 60 days. Second, Cluster I2 (or S2), which corresponds to Area X generated by the reversible jump MCMC, disappeared once the main earthquake occurred. Third, earthquakes in Cluster L2 (corresponding to M4 or E4 or Area Y in the spatial clustering results) appeared after the occurrence of the main earthquake. Both the second and the third points indicate that in the research area the tectonic

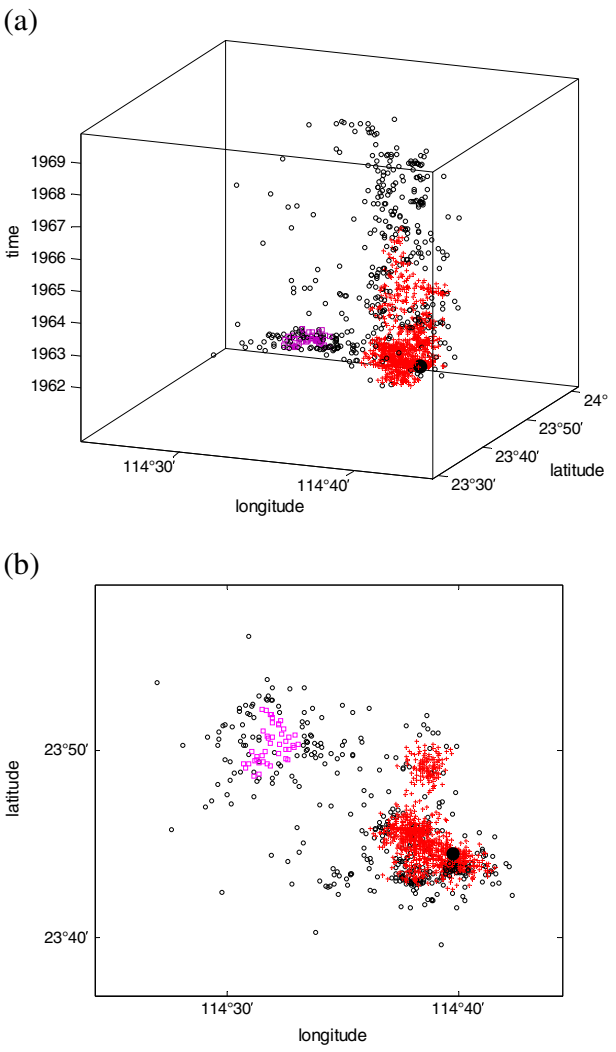


Fig. 17 Clustering results generated for a temporal window width = 1 year. 3D view; (**b**) 2D view (Cluster I1 is indicated by red crosses; Cluster I2 by purple rectangles; noise by black circles and the main earthquake by a larger black dots)

stress was redistributed after the main earthquake occurred. Fourth, we can explain why there are higher density clusters (i.e. M5 and M6) in Cluster M3. This is because both foreshocks and aftershocks were located in the support domain of M5 and M6. In other words, clustered earthquakes in this area never stopped during the whole period. Fifth, Clusters S4 and S5 were the aftershocks of two offspring strong earthquakes (in detail, one occurred on 8 November 1962, measuring 4.8 M; the other occurred on 23 September 1964, measuring 5.1 M). Previous research results have proved that this was the case [50, 53].

7 Discussion and conclusions

In this paper, we constructed a theory of multi-scale decomposition of point process data based on the *k*th nearest distance transformation. In the theory, point process data are treated

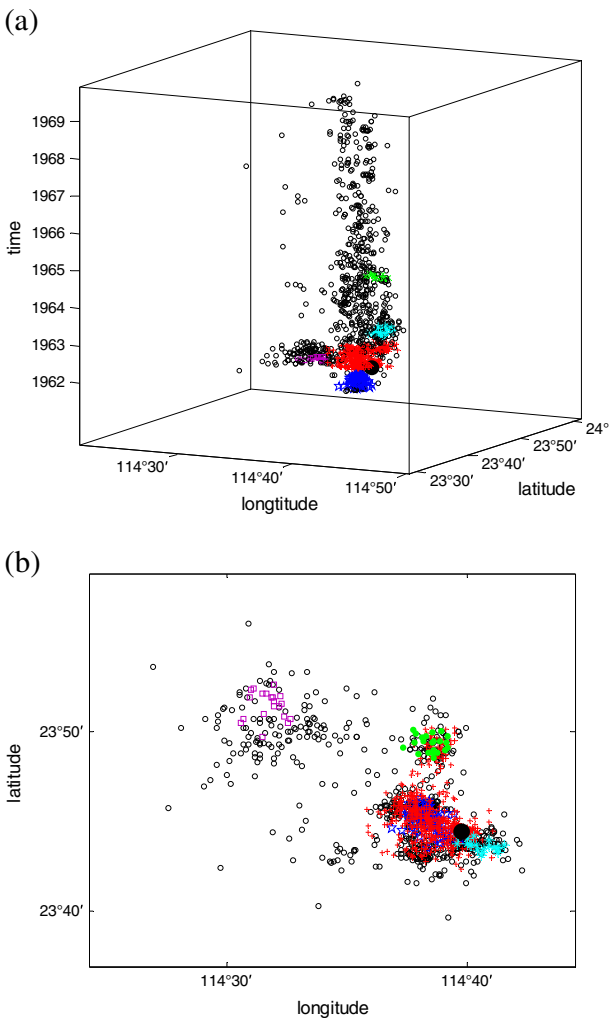


Fig. 18 Clustering results generated for a temporal window = 60 days. (**a**) 3D view; (**b**) 2D view (Cluster S1 is indicated by red crosses; Cluster S2 by purple rectangles; Cluster S3 by blue stars; Cluster S4 by cyan asterisks; Cluster S5 by green solid dots; noise by black circles and the main earthquake by a larger black dot)

as the summation of homogeneous point processes. The components (i.e. the homogeneous processes) are identified in the light of the probability density distribution of their *k*th nearest distances. The procedure for the separation is composed of two major steps. The first is to differentiate between point processes of multiple densities. To decompose point process data into homogeneous processes, three different strategies are employed according to different conditions (see Fig. 2). The second step is to identify distinct clusters from each process according to the spatial (or spatiotemporal) density connectivity of the points.

The contribution of the theory can be compared to that for wavelet transformation in signal processing. That is, with our theory, any complex point process data set can be separated into different homogeneous point processes while the wavelet transformation can decompose any function into wavelet components of different frequencies. Any identified homogeneous point process can be seen as a simple process with constant density over its support domain. It might represent a component of the point process data over a certain scale. The theory can also decompose the spatial-temporal data into different spatial-

temporal components with the windowed *k*th nearest distance. It can identify distinct spatiotemporal clusters over varied temporal scales or spatial scales as the temporal window width changes.

There are two limitations to this theory, which may lead to future work. That is, the theory cannot deal with the regular distributed point data because the *k*th nearest distances of points in a regular pattern are not subject to the pdf given by Eq. 4. Second, the decomposition of the point process data is a time consuming process. The complexity of the reversible jump MCMC is $O(T(Mk+(k\text{-}1)k!))$, where T is the iteration times, M is the number of points in the data. The complexity of stepwise EM is N^2MT, where N is the number of processes in the data. As a result, it needs a large amount of time when M and N are larger. More work needs to be done to reduce the computational time.

Acknowledgement This study was funded through support from the National Natural Science Foundation of China (Project Number: 41171345), a grant from the Chinese Academy of Sciences (KZCX2-YW-QN303), and grants from National Key Technology R&D Program (No. 2012AA12A403 and 2011BAH06B03).

References

1. Agrawal R, Gehrke J, Gunopulos D, Raghavan P (1998) Automatic subspace clustering of high dimensional data for data mining applications. In: Proceedings of the ACM SIGMOD '98 international conference on management of data. Seattle, WA, USA, pp. 94–105
2. Allard D, Fraley C (1997) Nonparametric maximun likelihood Estimation of features in spatial point process using voronoi tessellation. J Am Stat Assoc 92:1485–1493
3. Ankerst M, Breunig MM, Kriegel H-P, Sander J (1999) OPTICS: Ordering points to identify the clustering structure. In: Proceedings of ACM-SIGMOD'99 International Conference on Management Data (Philadelphia: PA), pp 46–60
4. Banfield JD, Raftery AE (1993) Model based Gaussian and non-Gaussian clustering. Biometrics 49:803–821
5. Besag JE, Gleaves JT (1973) On the detection of spatial pattern in plant communities. Bulletin of the International Statistical Institute 45(Book 1):153–158
6. Byers S, Raftery AE (1998) Nearest-neighbor clutter removal for estimating features in spatial point processes. J Am Stat Assoc 93:557–584
7. Byth K, Ripley BD (1980) On sampling spatial patterns by distance methods. Biometrics 36:279–284
8. China Meteorological Data Sharing Service System, http://cdc.cma.gov.cn, accessed on March 2011
9. Cressie NAC (1991) Statistics for spatial data (1st edition). John Wiley & Sons, Inc., New York (Chapter 8)
10. Daszykowski M, Walczak B, Massart DL (2001) Looking for natural patterns in data Part 1. Density-based approach. Chemometr Intell Lab Syst 56:83–92
11. Deng M, Liu QL, Cheng T, Shi Y (2011) An adaptive spatial clustering algorithm based on delaunay triangulation. Comput Environ Urban Syst 35:320–332
12. Diggle PJ (1985) A kernel method for smoothing point process data. Appl Stat 34:138–147
13. Duczmal L, Cançado ALF, Takahashi RHC, Bessegato LF (2007) A genetic algorithm for irregularly shaped spatial scan statistics. Comput Stat Data Anal 52:43–52
14. Eberhardt LL (1967) Some developments in 'distance sampling'. Biometrics 23:207–216
15. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with Noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, pp 226–231
16. Estivill-Castro V, Lee I (2002) Multi-level clustering and its visualization for exploratory spatial analysis. GeoInformatica 6:123–152
17. Estivill-Castro V, Lee I (2002) Argument free clustering for large spatial point data sets. Comput Environ Urban Syst 26:315–334
18. Filippone M, Sanguinetti G (2011) Approximate inference of the bandwidth in multivariate kernel density estimation. Comput Stat Data Anal 55:3104–3122

19. Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. Computer Journal 41:578–588
20. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 97:611–631
21. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I (2008) Space–time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. BMC Public Health 6:1–13
22. Han JW, Kamber M, Tung AKH (2001) Spatial clustering methods in data mining. In: Miller HJ, Han JW (eds) Geographic data mining and knowledge discovery. Taylor & Francis, London, pp 188–217
23. Hinneburg A, Keim DA (1998) An efficient approach to clustering in large multimedia databases with noise. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, USA, New York, pp. 58–65
24. Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. John Wiley & Sons Ltd: West Sussex, 517 p
25. Jin HD, Leung KS, Wong ML, Xu ZB (2005) Scalable model-based cluster analysis using clustering features. Pattern Recognition 38:637–649
26. Johnson RB, Zimmer WJ (1985) A more powerful test for dispersion using distance measurements. Ecology 6:1669–1675
27. Karypis G, Han EH, Kumar V (1999) Chameleon: hierarchical clustering using dynamic modelling. IEEE Computer 32:68–75
28. Kulldorff M, Nagarwalla N (1995) Spatial disease clusters: detection and Inference. Stat Med 14:799–810
29. Kulldorff M (1997) A spatial scan statistic. Comm Stat Theor Meth 26:1481–1496
30. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F (2005) A space–time permutation scan statistic for disease outbreak detection. PLoS Med 2:216–224
31. Lin C-R, Chen M-S (2005) Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging. IEEE Trans Knowl Data Eng 17:145–159
32. Lin CY, Chang CC (2005) A new density-based scheme for clustering based on genetic algorithm. Fundamenta Informaticae 68:315–331
33. Liu D, Nosovskiy GV, Sourina O (2008) Effective clustering and boundary detection algorithm based on Delaunay triangulation. Pattern Recogn Lett 29:1261–1273
34. Liu P, Zhou D, Wu NJ (2007) VDBSCAN: varied density based spatial clustering of applications with noise. In: Proceedings of IEEE international conference on service systems and service management, Chengdu, China, pp 1–4
35. Murtagh F, Starck JL (1998) Pattern clustering based on noise modeling in wavelet space. Pattern Recogn 31(7):847–855
36. Neill DB (2006) Detection of spatial and spatio-temporal clusters. Ph.D. Thesis of University of South Carolina
37. Neill DB, Moore AW (2005) Anomalous spatial cluster detection. In: Proceeding of KDD 2005 workshop on data mining methods for anomaly detection. Chicago, Illinois, USA, pp 41–44
38. Pascual D, Pla F, Sanchez JS (2006) Non parametric local density-based clustering for multimodal overlapping distributions. In: Proceedings of intelligent data engineering and automated learning (IDEAL2006), Spain, Burgos, pp 671–678
39. Pei T (2011) A non-parameter index for differentiating between heterogeneity and randomness. Math Geosci 43:345–362
40. Pei T, Jasra A, Hand DJ, Zhu AX (2009) DECODE: a new method for discovering clusters of different densities in spatial data. Data Min Knowl Discov 18:337–369
41. Pei T, Zhu AX, Zhou CH, Li BL, Qin CZ (2006) A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. Int J Geogr Inform Sci 20:153–168
42. Pei T, Zhu AX, Zhou CH, Li BL, Qin CZ (2009) Detecting feature from spatial point processes using Collective Nearest Neighbor. Comput Environ Urban Syst 33(6):435–447
43. Pei T, Zhou CH, Zhu AX, Li BL, Qin CZ (2010) Windowed nearest-neighbour method for mining spatio-temporal clusters in the presence of noise. Int J Geogr Inform Sci 24(6):925–948
44. Prayag VR, Deshmukh SR (2000) Testing randomness of spatial pattern using Eberhardt's index. Environmetrics 11:571–582
45. Ripley BD (1977) Modelling spatial patterns. J Roy Stat Soc B 39:172–192
46. Roy S, Bhattacharyya DK (2005) An approach to find embedded clusters using density based techniques. Lect Notes Comput Sci 3816:523–535
47. Sander J, Ester M, Kriegel H, Xu X (1998) Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. Data Min Knowl Discov 2:169–194
48. Schiffers K, Schurr FM, Tielborger K, Urbach C, Moloney K, Jeltsch F (2008) Dealing with virtual aggregation—a new index for analysing heterogeneous point patterns. Ecography 31:545–555

49. Sheikholeslami G, Chatterjee S, Zhang A (1998) WaveCluster: a multi-resolution clustering approach for very large spatial databases. In: Proceedings of the 24th international conference on very large data bases, New York City, NY, pp. 428–439
50. Shen CG, Chen HQ, Zhang HC, Huang LS, Li ZQ, Yang ZR, Wang DJ (1974) Xingfengjiang Reservoir Impounding earthquake and its influence on the Dam. Science In China 12(2):184–205
51. Tran TN, Wehrensa R, Lutgarde MCB (2006) KNN-kernel density-based clustering for high-dimensional multivariate data. Comput Stat Data Anal 51:513–525
52. Wan Y, Pei T, Zhou CH, Jiang Y, Qu CX, Qiao YL (2012) ACOMCD: a multiple cluster detection algorithm based on the spatial scan statistic and ant colony optimization. Comput Stat Data Anal 56:283–296
53. Wang MY, Yang MY, Hu YL, Li ZQ, Chen YT, Jin Y, Feng R (1976) Mechanism of the reservoir impounding earthquakes at Xinfengjiang and a preliminary endeavour to discuss their cause. Science in China 14(1):85–97
54. Yan P, Clayton MK (2006) A cluster model for space–time disease counts. Stat Med 25:867–881
55. Young YT, Chang LJ (2007) Bayesian nearest-neighbor analysis via record value statistics and nonhomogeneous spatial Poisson processes. Comput Stat Data Anal 51(9):4438–4449

Tao Pei obtained his B.S. in Geochemistry and Ph.D. in geostatistics from China University of Geosciences in 1993 and 1998 respectively. He has been an associate professor at Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences since July 2000. His research interest is in spatial data mining, in particular, cluster methods for spatiotemporal data. His other research interests include geostatistics and GIS.

Jianhuan Gao is pursuing his master degree at Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interest is the pattern mining in spatiotemporal sequences.

Ting Ma obtained his B.S. and M.S. from Peking University in 1999 and 2002 respectively and Ph.D. at Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, in 2006. He is now an assistant professor at Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include Parallel GIS developing and remote sensing of environment and ecology.

Chenghu Zhou obtained his B.S. from Nanjing University in 1984 and M.S. and Ph.D. at the Institute of Geography, Chinese Academy of Sciences in 1987 and 1992, respectively. He is now a professor and director of the Laboratory of Resources and Environmental Information System, Chinese Academy of Sciences. His research interests are in geographical information analysis and application models, remotely sensed image geo-understanding and analysis, spatial data mining and knowledge discovery.