

Article

# Decomposition of Repulsive Clusters in Complex Point Processes with Heterogeneous Components

Ci Song<sup>1,2</sup>  and Tao Pei<sup>1,2,\*</sup> 

<sup>1</sup> State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China

\* Correspondence: peit@reis.ac.cn; Tel.: +86-010-6488-8960

Received: 10 May 2019; Accepted: 24 July 2019; Published: 26 July 2019



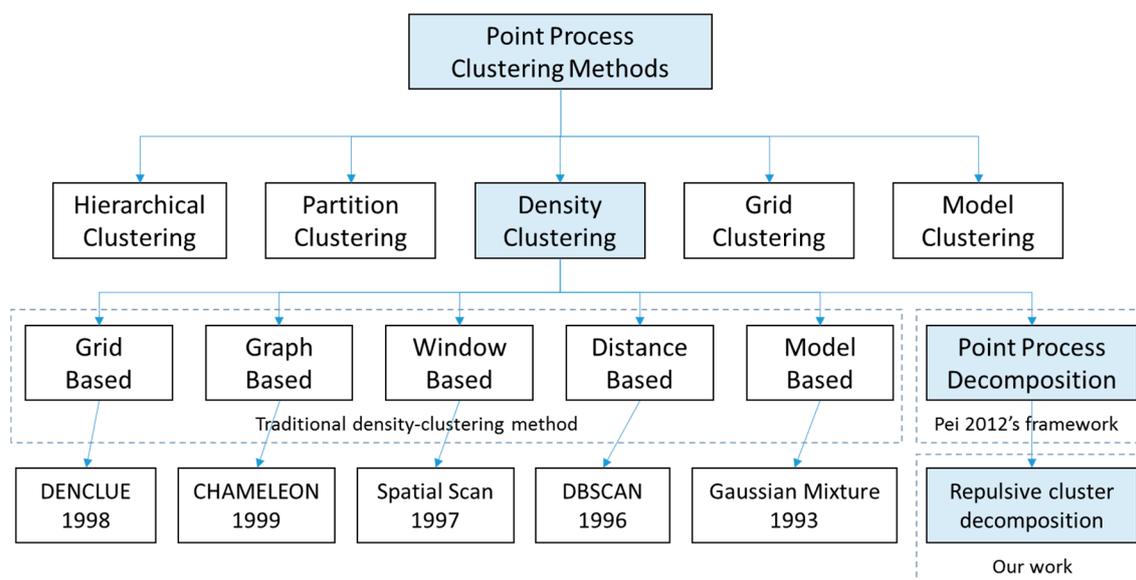
**Abstract:** The decomposition of a point process is useful for the analysis of spatial patterns and in the discovery of potential mechanisms of geographic phenomena. However, when a local repulsive cluster is present in a complex heterogeneous point process, the traditional solution, which is based on clustering, may be invalid for decomposition because a repulsive pattern is not subject to a specific probability distribution function and the effects of aggregative and repulsive components may be counterbalanced. To solve this problem, this paper proposes a method of decomposing repulsive clusters in complex point processes with multiple heterogeneous components. A repulsive cluster is defined as a set of repulsive density-connected points that are separated by a certain distance at a small scale and aggregated at a large scale simultaneously. The H-function is used to identify repulsive clusters by determining the repulsive distance and extracting repulsive points for further clustering. Through simulation experiments based on three datasets, the proposed method has been shown to effectively perform repulsive cluster decomposition in heterogeneous point processes. A case study of the point of interest (POI) dataset in Beijing also indicates that the method can identify meaningful repulsive clusters from types of POIs that represent different service characteristics of shops in different local regions.

**Keywords:** decomposition of a point process; spatial heterogeneity; repulsive cluster; aggregative cluster; H-function

## 1. Introduction

Many geographic phenomena affected by various factors at different scales can be seen as complex point processes with components of different types and spatiotemporal scales, such as seismic events [1], crime incidents [2], commercial site locations, and other phenomena. The decomposition of these point processes into different components over different scales may help identify the corresponding spatiotemporal patterns and mechanisms. For example, a seismic dataset can be decomposed as the superposition of background earthquakes at the global scale and clustered earthquakes at the local scale. Crime incidents can be decomposed into random cases distributed around a city (low-density noise), and high-risk cases concentrated in hotspots (high-density clusters). Chain stores in a large city can also be decomposed into several isolated shops (repulsive clusters) distributed in districts to serve local communities or many stores (aggregative clusters) concentrated in central business district (CBD) areas. These patterns are widely used for identifying areas with high earthquake rates, predicting periods, areas in a city with high-crime incidence [3], as well as providing valuable information for site selection and urban planning. Thus, the decomposition of point processes is a useful theoretical method in geographical research.

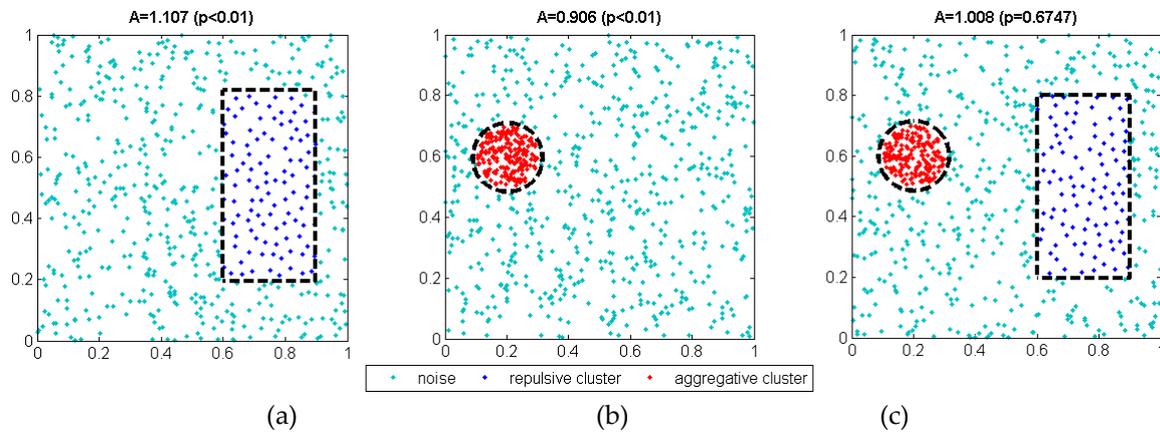
Recently, several methods based on various clustering algorithms have been developed to decompose the point processes of different geographical phenomena [4–6]. One commonly considered problem is the separation of homogeneous aggregative clusters with different densities from the point set. Many density-based clustering methods, such as DENCLUE [7], CHAMELEON [8], DBSCAN [9], OPTICS [10], among others, have been proposed to solve this problem. Pei [11] summarized these clustering methods and classified them into five types, including grid-based [12], graph-based [13], window-based [14], distance-based [15] and model-based models; the study also constructed a theoretical framework to objectively decompose clusters with multiple densities and arbitrary shapes in a complex point data set, as seen in Figure 1. Previous research has determined the heterogeneity of a point data set based on summary indices [16–20] or scale-related X-functions, such as the L-function and K-function [21,22]. Subsequently, a likelihood model of the  $k$ th nearest distance for a point process with multiple density components is constructed on the basis of various parameter estimation methods, such as the EM algorithm [1] and the reversible jump MCMC algorithm [15,23,24]. Finally, points are connected to clusters of different densities based on the connectivity at different scales through the parameter  $\epsilon$ .



**Figure 1.** Taxonomy of point process clustering methods.

Most of the clustering methods mentioned above have certain limitations in real data applications because they are based on the hypothesis that the point set is only composed of homogeneous aggregative clusters and noise; accordingly, these clustering methods do not consider repulsive clusters. Repulsive clusters are ubiquitous in geographic phenomena and are commonly used to describe the competition among plants [25,26], animals [27], the distribution of aerial base stations [28], and the service areas of chain stores [29], among other applications [30–32]. When a local repulsive cluster is present in a complex heterogeneous point process (with multiple heterogeneous components), the commonly used statistical measurements for quantifying heterogeneity may be invalid in some situations. For example, as seen in Figure 2, although the statistical measurement A [33,34] effectively quantifies the heterogeneity of two point processes, as seen in Figure 2a,b, that contain homogeneous components, it cannot effectively quantify the heterogeneity of a complex point process with heterogeneous components, as seen in Figure 2c with one repulsive cluster and one aggregative cluster; In this case, an incorrect CSR (complete spatial randomness) distribution judgment may be made. Specifically, a repulsive pattern is not subject to the same specific probability distribution function as homogeneous components in a point process. Additionally, the effects of the aggregative and repulsive components may be counterbalanced, which can lead to an indeterminate

statistical result. Although these statistical measurements are significant ( $A < 1$  or  $A > 1$ ), they can only indicate one aspect of a certain component (repulsion or aggregation) for the point set, and it is difficult to understand the spatial heterogeneity of a complex point process with multiple heterogeneous components on the basis of these values.



**Figure 2.** Comparison of point processes with different components. (a) Noise and a repulsive cluster; (b) noise and an aggregative cluster; and (c) noise, a repulsive cluster and an aggregative cluster.

To solve this problem, we propose a method of decomposing repulsive clusters from complex point processes with multiple heterogeneous components to improve the theoretical method of point process decomposition. In our study, a repulsive cluster is considered to be a connected dense region that consists of repulsive points with respect to a certain distance. This definition differs from that of a traditional aggregative cluster in which points are separated by a certain distance at a small scale. Additionally, the repulsive clusters still exhibit a certain aggregate morphology at a large scale and noise is isolated.

The remainder of the paper is arranged as follows. Several concepts regarding repulsive clusters and details of the method for decomposition of repulsive clusters are introduced in Section 2. Three groups of simulation experiments and related parameter analysis are performed to illustrate the effectiveness of the method in Section 3. A case study of different types of point of interest (POI) distributions in Beijing is presented in Section 4. The conclusions and future work are detailed in Section 5.

## 2. Materials and Methods

### 2.1. Basic Concepts

Before our method is proposed, some basic concepts related to repulsive clusters should be introduced in Figure 3.

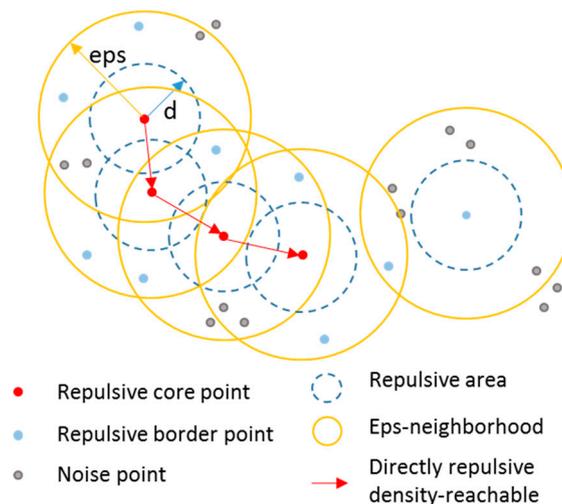
**Definition 1** (Repulsive point). Let  $D$  be a set of points. A point  $p$  is a repulsive point with respect to distance  $d$ . If  $\forall q \in D, q \neq p \Rightarrow \text{dist}(p, q) \geq d$ , then this repulsive point  $p$  can be denoted as  $p^d$  and the repulsive point set can be denoted as  $D^d$ . The area within distance  $d$  of each repulsive point  $p$  is denoted as the repulsive area.

**Definition 2** (Repulsive  $\text{eps}$  neighborhood and core repulsive point). The repulsive  $\text{eps}$  neighborhood of a repulsive point  $p^d$ , denoted as  $N_{\text{eps}}(p^d)$ , is defined as  $N_{\text{eps}}(p^d) = \{q | \text{dist}(p, q) \leq \text{eps}, q \in D^d\}$ , where  $p \in D^d, \text{eps} > d$ . If  $|N_{\text{eps}}(p^d)| \geq \text{minpt}$ ,  $p^d$  is a core repulsive point with respect to  $d$  and  $\text{minpt}$ .

**Definition 3** (Directly repulsive density-reachable). If  $q^d \in N_{\text{eps}}(p^d)$  and  $p^d$  is a core repulsive point with respect to  $d$  and  $\text{minpt}$ , then  $q^d$  is directly repulsive density-reachable from a core repulsive point  $p^d$ .

**Definition 4** (Repulsive density-connected). Repulsive point  $o$  is considered to be repulsive density-connected to point  $p$  if there is a collection of repulsive core points  $q_1^d, q_2^d, \dots, q_n^d$ , where  $o = q_0^d, p = q_n^d$ , such that  $q_{i-1}^d$  is directly repulsive density-reachable from  $q_i^d, i = 1, 2, \dots, n$ .

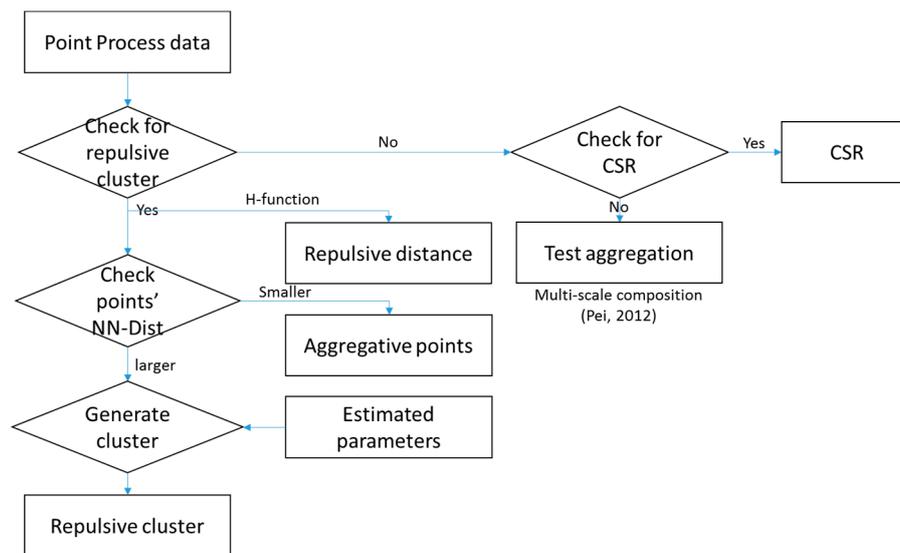
**Definition 5** (Repulsive cluster). A repulsive cluster  $C^d$  is a non-empty subset of  $D$  that satisfies the following three conditions. (i)  $\forall p \in C^d$ , where  $p$  is a repulsive point with respect to  $d$ . (ii)  $\forall p, q \in D$ , where if  $p \in C^d$  and  $p$  is repulsive density connected to  $q$ , then  $q \in C^d$ . (iii)  $\forall p, q \in C^d$ , where  $p$  is repulsive density connected to  $q$ .



**Figure 3.** Basic concepts relating to repulsive clusters.

## 2.2. Method

On the basis of the concepts defined above, we propose a method to decompose the repulsive clusters from a point process with multiple heterogeneous components. This method can be divided into three steps, as shown in Figure 4. First, we determine whether there is a repulsive cluster in the heterogeneous data set using the H-function. If there is a repulsive cluster, we proceed to the second step. Otherwise, the point set is considered to have no repulsive clusters and can be tested for aggregative clusters based on Pei's theory [11]. Second, if a repulsive cluster is confirmed, a repulsive distance with respect to the repulsive cluster is determined and used to eliminate certain local aggregative points from the data set. The remaining points only include repulsive clusters to be identified and noise. Third, we determine  $\epsilon$ , which is used to construct the density domain of repulsive points and distinguish all repulsive core points from noise. On the basis of this approach, we can generate repulsive clusters according to the density connectivity of those points. The details of this method are as follows.



**Figure 4.** Process of decomposing repulsive clusters from the point process.

### 2.2.1. Determining the Existence of Repulsive Clusters

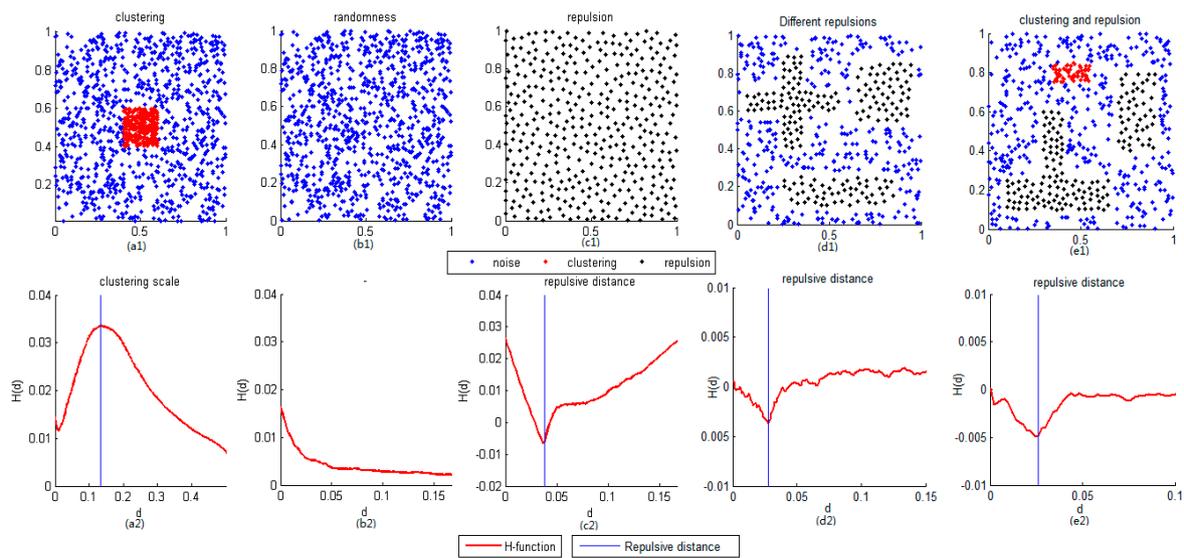
Because traditional indices have certain limitations in indicating the distribution of complex heterogeneous point processes, we use an interactive method to determine whether there is a repulsive cluster in the dataset. Previous studies have shown that Ripley's K-function and H-function [35,36] can be widely used to compare a simple point distribution with a random distribution. For a simple point process, a positive value of  $H(d)$  indicates clustering over the given spatial scale, whereas a negative value indicates dispersion [37]. Therefore, we can use the H-function for dispersion identification. The definitions of the H-function and K-function are as follows:

$$K(d) = \lambda^{-1} \sum_{i=1}^n \sum_{j=1}^n \delta_{i,j}(d) / n, (i, j = 1, 2, \dots, n; i \neq j) \quad (1)$$

$$H(d) = \sqrt{K(d) / \pi} - d \quad (2)$$

where  $\delta_{ij}(d) = \begin{cases} 1, & (d_{ij} \leq d) \\ 0, & (d_{ij} > d) \end{cases}$ ,  $d_{ij}$  is the distance between point  $i$  and point  $j$ ,  $n$  is the number of points, and  $\lambda$  is the density of points in the study area.

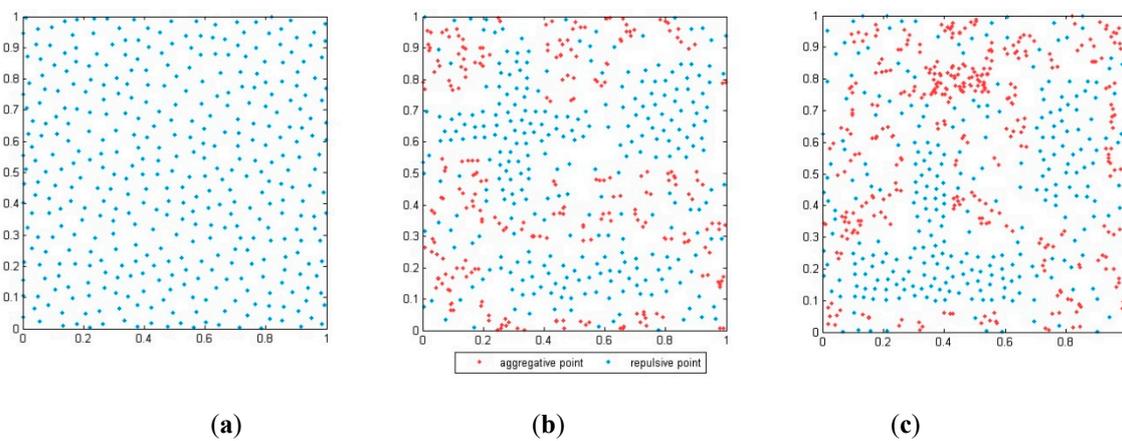
Because a significant local peak value in the H-function reflects the cluster scale of the aggregative pattern in the point process [38], the apparent local valley value in the H-function is believed to be an indicator of a repulsive cluster with respect to a certain repulsive distance. Here, we use a minima detection algorithm [39] to determine whether there are repulsive clusters by identifying the local valley in the H-function. This algorithm is to detect the first derivative of a peak/valley, which has a downward-going/upward-going zero-crossing at the peak maximum/valley minimum. Thus, the local valley position in the H-function can be determined by the upward-going zero-crossing of the maximum slope exceeding a certain threshold (empirically set to 0.001). If there is a significant repulsive cluster in the point process, a significant local valley can be identified in the H-function, as seen in Figure 5c,d,e. Otherwise, no significant local valley will be present in the H-function plot, as seen in Figure 5a,b. The specific scale  $d$  corresponding to an identified local valley of the H-function is considered to be the repulsive distance because the number of neighbors  $N_{eps}(p^d)$  changes most when  $d$  is close to the repulsive distance. This change will manifest itself as a local valley in the H-function. Thus, the existence of repulsive clusters can be checked.



**Figure 5.** Distributions of different types of point processes and their H-functions. (a) Aggregation and noise; (b) homogeneous point process; (c) repulsion (repulsive distance = 0.038); (d) repulsion and noise (repulsive distance = 0.028); and (e) aggregation, repulsion and noise (repulsive distance = 0.026).

### 2.2.2. Extracting Repulsive Points Based on the Repulsive Distance

Once the existence of a repulsive cluster is confirmed and the repulsive distance is determined, the nearest distance (*NN-Dist*) of each point in the dataset is compared with the repulsive distance and all points with an *NN-Dist* less than the repulsive distance are decomposed as aggregative components, as seen in Figure 6b,c, red points. The remaining points containing repulsive clusters and noise are processed in the next step.



**Figure 6.** Extraction of repulsive points. Blue points represent repulsive points with respect to  $d$ , and red points represent aggregative components. (a) repulsion; (b) repulsion and noise; (c) aggregation, repulsion and noise.

Note that the repulsive distance will be limited to a certain range ( $0 \sim 1.0746\lambda^{-1/2}$  [33]) depending on the size and density of the repulsive cluster. Larger estimated repulsive distances correspond to a lower amount of retained noise but a greater likelihood that positive repulsive point decomposition will occur and vice versa. The effect of the repulsive distance estimation on the cluster result will be detailed in Section 3.2.1.

### 2.2.3. Generating Density-Connected Repulsive Clusters

After the extraction process, we decomposed all aggregative components of the point set and obtained the remaining repulsive points with respect to the repulsive distance. Thus, the following steps were used to estimate  $\epsilon$ , determine the cluster scale and generate density-connected repulsive clusters.

Because  $\epsilon$  is the only parameter required to transform spatial points into the density domain, it is very important to estimate the corresponding ability to separate clusters from noise. To estimate  $\epsilon$ , we use the same methods as Pei [38] and determine the appropriate value of the clustering scale factor based on Monte Carlo simulation experiments. The difference is that we use the local maximum of the H-function to estimate the clustering scale instead of the minimum of the derivative function because the derivative function of a repulsive point set tends to fluctuate. As  $\epsilon$  is the product of the clustering scale and clustering scale factor, the experiments based on synthetic data indicated that when the clustering scale factor is approximately 5/12, the identification error is minimized, as seen in Section 3.2.2.

After  $\epsilon$  is determined, the repulsive clusters can be generated in the corresponding regions of the density domain. All core repulsive points are marked and connected to each other based on the density connectivity according to definition 2.4. The detailed process is as follows. First, we randomly choose a core repulsive point and assign a cluster ID. Then, all neighbors of this point are assigned the same cluster ID and traversed to expand the cluster region. The same process is repeated for the next core repulsive point in the neighborhood until no unclassified core points are found in the neighborhood. All points with the same cluster ID form a cluster, and another cluster begins with a new unclassified core repulsive point. After all core repulsive points are traversed, all repulsive clusters are generated from the data set.

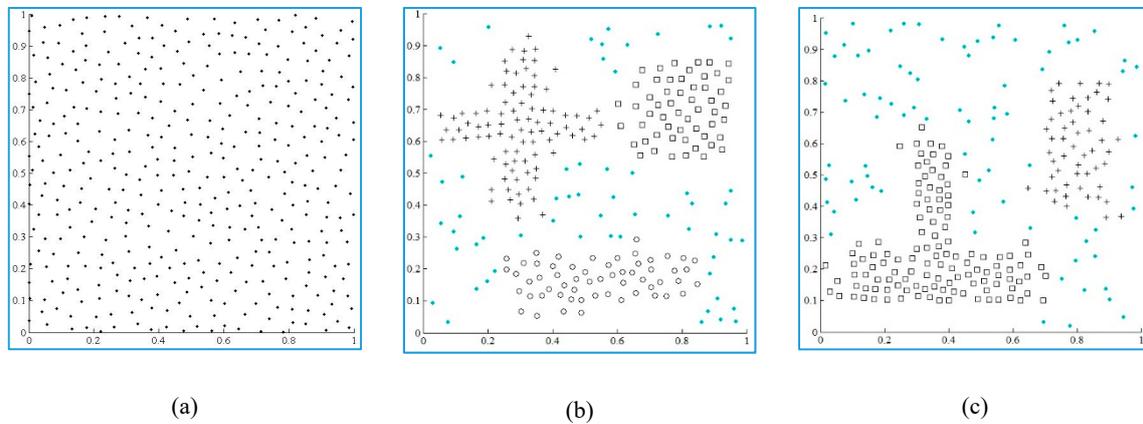
## 3. Simulation Results

### 3.1. Validation of the Algorithm for Different Synthetic Datasets

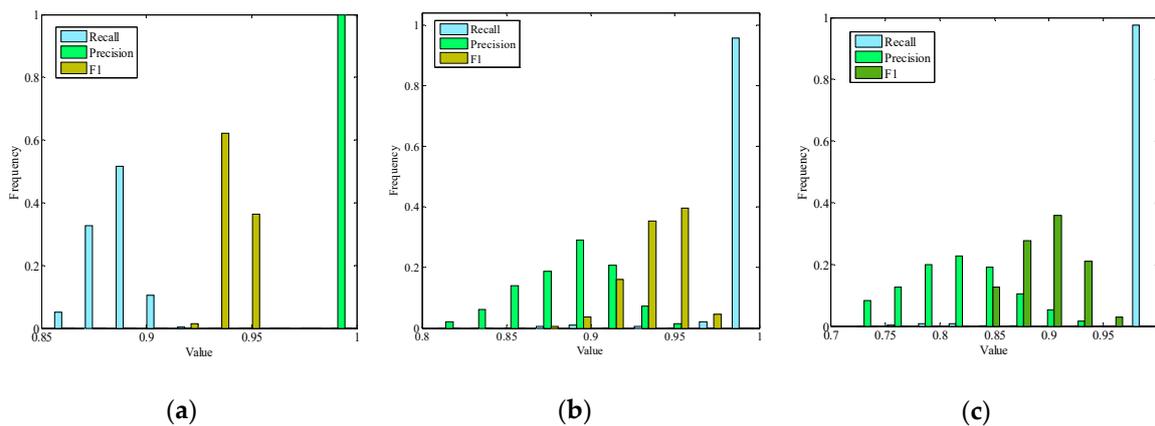
In this section, three groups of synthetic data in Figure 5c1, d1 and e1 are used to verify the proposed method. The Group I dataset, as seen in Figure 5c1, exhibits a global repulsive pattern and is entirely composed of repulsive points. The Group II dataset, as seen in Figure 5d1, is a complex point process that includes three repulsive clusters of different shapes (cross-shaped, square-shaped, and strip-shaped clusters) and densities (densities 1, 1, and 1.5 times that of the background noise). The Group III dataset, as seen in Figure 5e1, is a complex point process with heterogeneous components and includes two repulsive clusters (a reverse “T” shape with the same density as the background noise and a bar shape with 1.5 times the density of the background noise) and an aggregative cluster (a strip-shaped cluster with five times the density of the background noise). Experiments are performed for each dataset 1000 times to evaluate the average identification rates.

The identification result is shown in Figure 7 and the identification rate is shown in Figure 8. The simulation experiments for Group I data have a recall rate above 85% and a precision rate equal to 100%. Simulation experiments with the other two datasets yielded satisfactory recall rates (R above 0.99 in 95% of simulations) and precision rates (lowest rates of P are 0.8 and 0.73). F1 is defined as  $F_1 = 2RP/(R + P)$  and is commonly used to evaluate summary results. The F1 values obtained were above 0.9 in almost all simulations. The average detection results are listed in Table 1. This table indicates that the proposed method successfully identifies each repulsive cluster at a high rate. For Group I data, the recall rate is approximately 88.6% and the precision rate is 100% because almost all points have been correctly identified except a few points in edge regions. For Group II data, the recall rates of most simulations reach 100% and the precision rates are almost above 85% except for those of the “cross-shaped” cluster. The nearly perfect recall rates indicate that almost all points belonging to the repulsive clusters were identified by the proposed method. Additionally, the relatively low precision rates suggest that certain noise points were misidentified, including points around the borders of clusters, especially in the upper area of the “cross-shaped” cluster and in the lower-right

area of the “strip-shaped” cluster, as seen in Figure 7b. For Group III data, a slightly inferior result can be observed. Notably, the detection precision for each repulsive cluster decreases by approximately 7%. Although the recall rates remain high, the lowest precision rate decreases to approximately 73%. These results indicate that the detection may be disturbed to some extent when there is an aggregative cluster.



**Figure 7.** Generating density-connected repulsive clusters. (a) All points in this set have been identified as repulsive points and generated as a single repulsive cluster. (b) Three repulsive clusters have been identified, including cross, square and circle clusters. (c) Two repulsive clusters have been identified, including cross and square clusters.



**Figure 8.** Detection results from the simulations. (a) repulsion; (b) repulsion and noise; (c) aggregation, repulsion and noise.

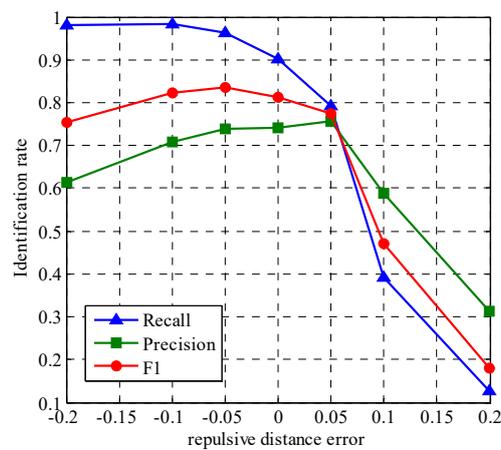
**Table 1.** Detection results for synthetic data.

| Dataset   | Clusters                    | Number of Points | TP     | FP    | FN    | Recall | Precision |
|-----------|-----------------------------|------------------|--------|-------|-------|--------|-----------|
| Group I   | All points                  | 400.5            | 354.76 | 0     | 45.71 | 88.6%  | 100%      |
|           | Square-Shaped Cluster       | 53.94            | 53.67  | 13.35 | 0.28  | 99.5%  | 88.2%     |
|           | Strip-Shaped Cluster        | 53.99            | 52.92  | 7.01  | 1.08  | 98.0%  | 89.6%     |
| Group II  | Cross-Shaped Cluster        | 80.98            | 80.57  | 19.18 | 0.41  | 99.5%  | 83.1%     |
|           | Noise Cluster               | 0.00             | 0.00   | 2.93  | 0.00  | 0.0%   | 0.0%      |
|           | All                         | 188.92           | 187.80 | 24.00 | 2.24  | 99.4%  | 88.8%     |
| Group III | Bar-Square                  | 48.07            | 47.33  | 12.04 | 0.74  | 98.4%  | 83.2%     |
|           | Reversed “T”-Shaped Cluster | 105.37           | 100.73 | 20.00 | 4.64  | 95.4%  | 84.0%     |
|           | Noise Cluster               | 0.00             | 0.00   | 8.43  | 0.00  | 0.0%   | 0.0%      |
|           | All                         | 153.63           | 148.96 | 35.55 | 9.33  | 96.9%  | 81.2%     |

### 3.2. Parameter Analysis

#### 3.2.1. Effect of the Repulsive Distance on the Clustering Results

As discussed in Section 2.2.3, the determination of the repulsive distance influences the results of cluster identification. Here, we generate 100 duplicates of the Group III dataset and implement the proposed method for each dataset using different repulsive distance, with errors ranging from -20% to 20% of true value. The average identification rates for different estimation errors of the repulsive distance are shown in Figure 9. This figure shows that the identification rate remains stable at a relatively high level when the estimated repulsive distance error is negative and displays a sharp decline when the estimated repulsive distance error is positive. A 20% negative error yields an F1 value above 0.75, whereas a 10% positive error leads to a poor F1 below 0.5.



**Figure 9.** Identification rate for different estimation errors of the repulsive distance.

The reason for this result may be that, although repulsive noise will increase considerably for an underestimated repulsive distance, the proposed generation process can effectively distinguish clusters from noise, especially with an appropriate eps. However, an overestimated repulsive distance may abruptly eliminate the real repulsive points, which can directly lead to poor identification performance. Therefore, an underestimated repulsive distance may yield a better result than an overestimated distance.

#### 3.2.2. Effect of the Determination of Eps on the Clustering Results

After the repulsive distance has been determined, eps is the key parameter used to obtain the clustering results. In this section, we generate 100 simulated datasets and implement the proposed method for each dataset using different clustering scale factors ranging from 8/24 to 16/24. The average identification rates are shown in Figure 10. Notably, as the clustering scale factor increases, the recall rate slowly increases and the precision rate decreases. The clustering scale factor corresponding to the optimal result is approximately 5/12. When the clustering scale factor is below 5/12, the recall rate is still above 0.7, and the precision rate is above 0.85. When the clustering scale factor is above 5/12, the recall rate increases to above 0.9, but the precision rate decreases to below 0.75. However, the F1 measure may maintain an acceptable level with an identification rate above 0.8 if the clustering scale factor ranges from 8/24 to 14/24, which indicates fair performance.

The effect of the clustering scale factor on the clustering results can be interpreted as follows. When the clustering scale factor is too large, more noisy repulsive points are misidentified with a large eps neighborhood, which may lead to a high recall rate and low precision rate. In contrast, when the clustering scale is too small, the repulsive clusters may be broken into fractions because the repulsive core points are minimally connected to each other due to the unstable local density of repulsive points

with an inhomogeneous distribution. In our case study, a carefully selected value of 5/12 was used to detect repulsive clusters.

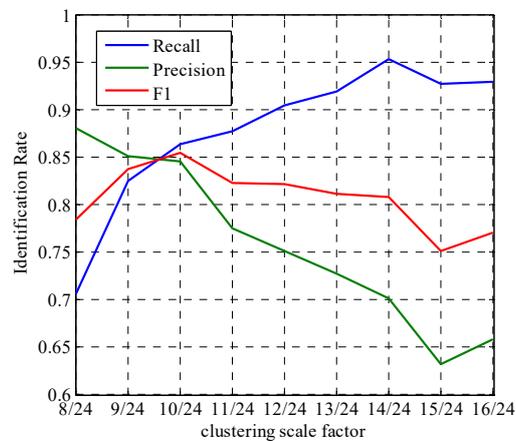


Figure 10. Identification rate with different clustering scale factors.

#### 4. Case Study and Discussion

To evaluate the proposed method based on a real dataset, we applied it to the POI dataset in Beijing to decompose different types of implicit repulsive clusters. This dataset provides different examples of POIs in urban areas that may reflect the demand situation and supply capacity of different POI services. Here, we chose eight types of POIs, including 7-Elevens, gas stations, KFC, kindergartens, McDonalds, shopping malls, Starbucks, and parks in Beijing, from the POI dataset in 2018. These types of POIs cover urban places where many of the daily activities of residents, such as eating, traveling, education, leisure, and entertainment activities, occur.

To avoid disturbances from small-scale POIs of the same types, we preprocessed the POI dataset by filtering out certain small POIs, such as the dessert shops of McDonalds and KFC. Kindergartens, shopping malls, and parks with areas less than certain thresholds were also removed. Here, we only analyze POIs within the fifth ring, which encircles the main built-up area in Beijing. After these steps, the POI dataset to be analyzed was obtained, as shown in Figure 11.

The H-function is shown in Figure 12 and was used to determine the existence of repulsive clusters for each type of POI. The figure shows that three types of POIs, i.e., KFC, McDonalds, and shopping malls, exhibit repulsive patterns, as seen in Figure 12 (solid lines). However, 7-Elevens and Starbucks exhibit aggregative patterns with increasing H-function trends, as seen in Figure 12 (dashed lines). Other types of POIs do not exhibit significant patterns in the corresponding H-functions (dashed-dotted lines). We also calculated the A index value for each type of POI to provide a comprehensive summary of the distribution patterns of POIs, as seen in Table 2. Table 2 shows that the H-functions and A indices both exhibit aggregative clustering patterns for 7-Eleven and Starbucks POIs. Although the A index values suggest an aggregative pattern for kindergartens, the H-function does not exhibit an upward trend at a large scale because the aggregative scale of kindergartens is too small to be identified in the H-function at a large scale due to the relatively high density of POIs in the area. The H-functions also indicate certain repulsive clustering patterns for KFC, McDonalds, and shopping mall POIs, whereas the A indices reflect aggregative clustering patterns for KFC and shopping malls. Thus, mixed patterns are observed for KFC and shopping mall POIs and a repulsive pattern is observed for McDonalds.

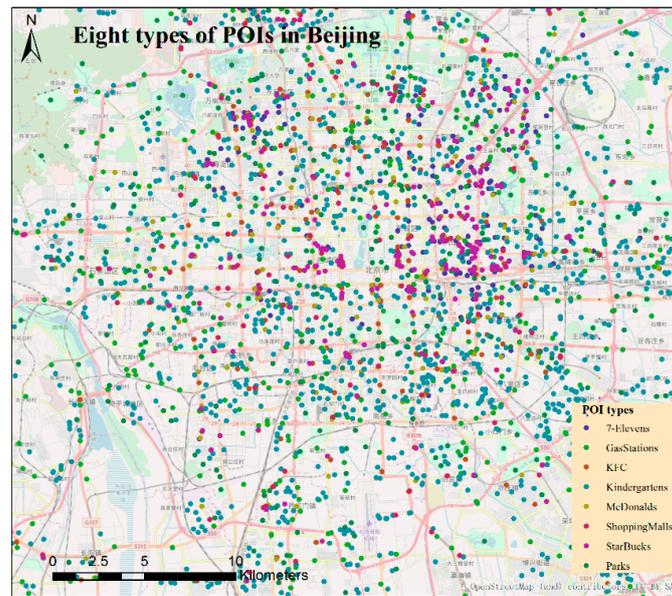


Figure 11. Distribution of different types of POIs in Beijing.

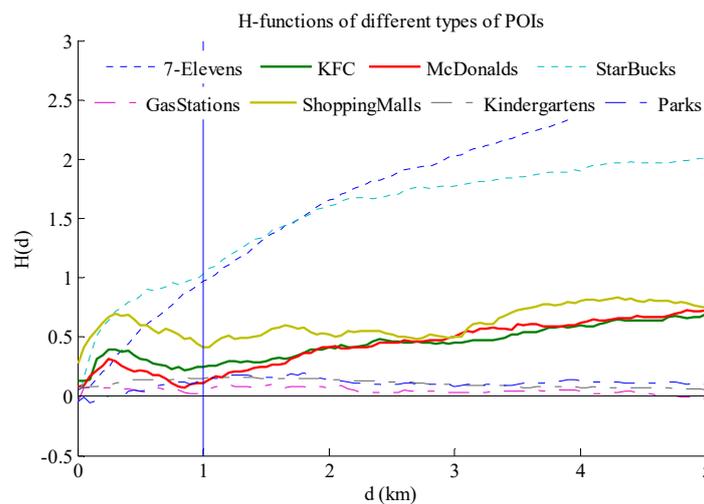


Figure 12. H-functions of different types of POIs. The aggregative clustering patterns shown are for 7-Elevens and Starbucks. The repulsive clustering patterns shown are for KFC, McDonalds, and shopping malls.

Here, we focus on types of POIs with repulsive clustering patterns (KFC, McDonalds, and shopping mall POIs). Because the local valleys of these three types of POIs are very close, the unified repulsive distance is set to 1 km for all three types of POIs for comparison. Similarly, a unified value of eps for detecting clusters is set to 2.1 km based on the method detailed in Section 2.2.3. The identified results are as follows.

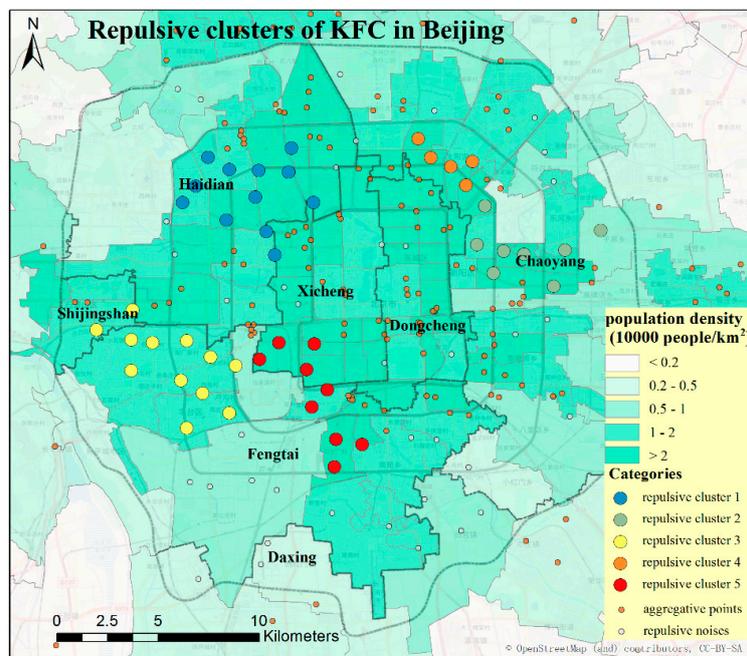
Figure 13a shows the KFC distribution in the fifth ring of Beijing and the population density of each background block. This figure shows three types of shops in the study area that represent repulsive clusters, aggregative points, and repulsive noise. Five repulsive clusters can be seen in the study area, each of which serves a specific area. These clusters are mainly distributed in densely populated areas. The blue cluster distributed in the fourth ring in the northwest serves several blocks in Haidian District, including the Haidian block, Shuguang block, Zizhuyuan block, Beixiaguan block, Exhibition Road block, among others. The orange cluster and green cluster are distributed in the eastern and northeastern areas, respectively, serve blocks in the Chaoyang District within the fifth ring. The orange

cluster covers several blocks, including the Xiaoguan block, Hepingjie block, and Xiangheyuan block, and the green cluster covers the Sanlitun block, Hujialou block, Tuanjiehu block, Liulitun block, among others. These two clusters are separated by Jingmi Road, which is the main road leading to the airport. The yellow cluster distributed between the fourth ring and the fifth ring in the southwest serve parts of Fengtai District and Shijingshan District in the fifth ring, including the Lugouqiao block, Fengtai block, Babaoshan block, and Lugu block. The red cluster distributed between the second ring and fourth ring mainly serves blocks in parts of Dongcheng District and Fengtai District, including the Guanganmen outer block, Baizhifang block, Youanmen block and Majiapu block. The aggregative shops are mainly distributed within the second ring and in population hotspot areas, such as Zhongguancun, the Beijing West Railway Station area, and the Beijing South Railway Station area. This pattern may be observed because many places within the second ring restrict development and shops can only be located in certain areas. Additionally, transportation hubs, such as the Beijing West Railway Station and Beijing South Railway Station, need more shops to meet the needs of a large number of people. Other shops associated with repulsive noise are mainly distributed outside the southern fourth ring, where the population is relatively small.

**Table 2.** Distribution pattern for different types of POIs.

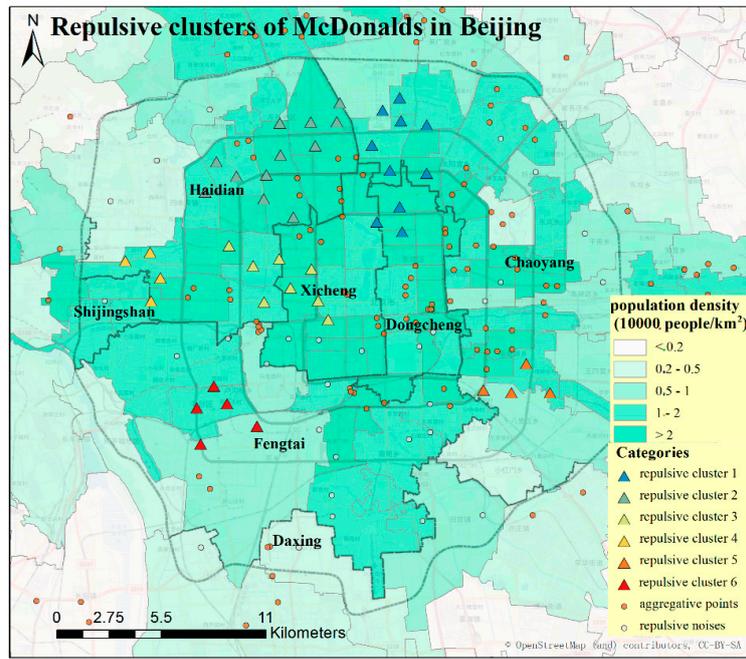
| POI Types      | Number of Points | H-function  | Clark and Evan’s A | Clustering Patterns |
|----------------|------------------|-------------|--------------------|---------------------|
| 7-Elevens      | 195              | uptrend     | 0.63 ***           | Aggregative         |
| KFC            | 200              | d = 0.9 km  | 0.94 *             | Mixed               |
| McDonalds      | 156              | d = 1 km    | 0.99               | Repulsive           |
| Starbucks      | 191              | uptrend     | 0.62 ***           | Aggregative         |
| Gas Stations   | 260              | -           | 1.04               | -                   |
| Shopping Malls | 128              | d = 1.05 km | 0.80 ***           | Mixed               |
| Kindergartens  | 830              | -           | 0.88 ***           | Aggregative         |
| Parks          | 216              | -           | 0.99               | -                   |

\*\*\* Significant at 0.001; \*\* significant at 0.01; and \* significant at 0.1.

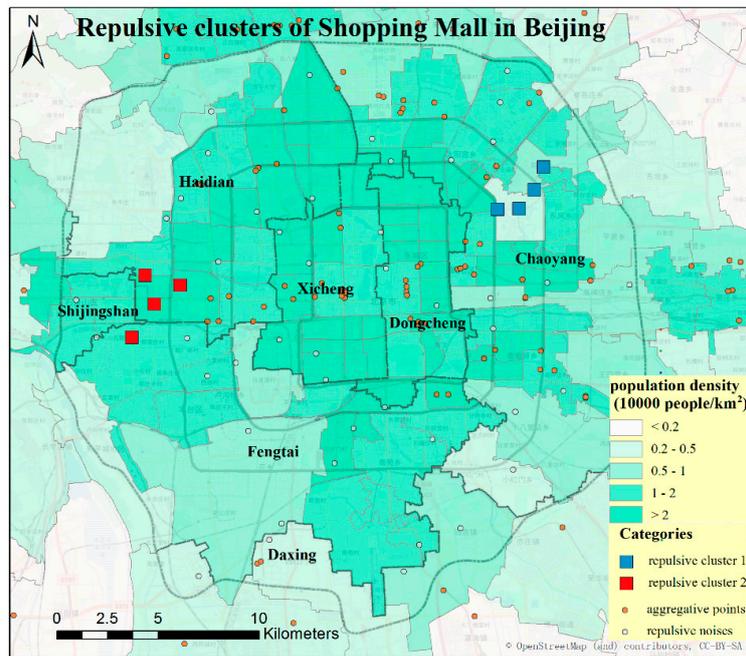


(a)

**Figure 13.** Cont.



(b)



(c)

**Figure 13.** Repulsive clusters of three types of point of interest (POIs) in Beijing. (a) Distribution of KFC; (b) distribution of McDonalds; and (c) distribution of shopping malls.

The McDonalds distribution is similar to that of KFC, as seen in Figure 13b. There are six repulsive clusters distributed in the study area, each of which serves several densely populated areas. Shops in the northern part of the study area include two repulsive green clusters that serve the northeastern portion of Haidian District and blue clusters that serve part of Chaoyang District. The yellow cluster and orange cluster distributed in the west third ring and west fifth ring, respectively, serve blocks in

the southern part of Haidian District and western part of Xicheng District, including the Tiancun Road block, Yongding Road block, Ganjiakou block, Yangfangdian block, Yuetan block and Exhibition Road block. The remaining two repulsive clusters of McDonalds shops are distributed in the southwestern part of the fourth ring in Fengtai District and the southeastern part of the fourth ring in Chaoyang District. These two clusters serve relatively small populations and areas. The aggregative McDonalds shops are distributed in Dongcheng District and at the same traffic hubs as KFC. In addition, there are many aggregative McDonalds shops distributed in Sanlitun and Chaoyang Park in the eastern third ring. The reason for this pattern may be similar to that for KFC. The repulsive noise points are dispensed in the fringe areas of the fifth ring, such as in Sijiqing Town in the northeastern fifth ring, the Nanyuan block and Dahongmen block in the southern fifth ring and the Pingfang block and Gaobeidian block in the eastern fifth ring.

The distribution of shopping malls exhibits a different pattern, and most of the shops exhibit an aggregative pattern. Figure 13c shows only two repulsive clusters distributed in the study area. One cluster is around the west third ring and serves the Yongding Road block and Wanshou Road block; the other is distributed in the northeast fourth ring and serves the Maizidian block. The aggregative shops are dominant in the central area, including the Zhongguancun block, Olympic Village block, Datun block, Jinrong Street block, Donghuamen block, and Chaowai block. The remaining repulsive noise shops are dispersed in the southern part of the study area, largely because the main business districts are concentrated in the northern and central parts of Beijing, and many sellers are centrally located in prime locations for competition. In this situation, the repulsive clusters of shopping malls identified in the study serve local residents instead of dominating the market.

## 5. Conclusions

In this study, a method is proposed to identify repulsive clusters in complex point processes and solve the decomposition problem in complex point processes with different heterogeneous components, including both repulsive clusters and aggregative clusters. Repulsive clusters, which are not considered in traditional point processes, consist of repulsive points that are separated by a certain distance at small scales and aggregated at large scales simultaneously. To illustrate the validity of the proposed method, the approach was tested in three simulation experiments and applied to identify the repulsive patterns of different types of POIs in Beijing. The identification rate in the simulation experiments reflected a satisfactory result, and the repulsive clusters of KFC, McDonalds, and shopping mall POIs also exhibited meaningful results that represent the service characteristics of shops in different regions.

The proposed framework provides a new point anomaly method for point processes and a solution for the decomposition of complex point processes with heterogeneous components. This approach may be useful in many applications for analyzing complex geographical phenomena. However, the method has some limitations. Notably, the repulsive clusters to be identified must be significant, which means that the size of a cluster should be sufficiently large and the repulsive distance should be sufficiently long. Otherwise, the H-function cannot accurately indicate the repulsive distance or there may be too much noise remaining to identify a target precisely. Future work may explore additional spatial characteristics of repulsive clusters and new indices to determine the existence of less significant repulsive clusters. Moreover, the method should be expanded to different applications, such as the identification of local service networks of public resources or the discovery of the local characteristics of competition.

**Author Contributions:** Conceptualization, Tao Pei and Ci Song; Methodology, Tao Pei and Ci Song; Software, Ci Song; Validation, Ci Song; Formal Analysis, Ci Song and Tao Pei; Investigation, Ci Song and Tao Pei; Data Curation, Ci Song; Writing-Original Draft Preparation, Ci Song; Writing-Review & Editing, Tao Pei and Ci Song; Visualization, Ci Song; Funding Acquisition, Tao Pei.

**Funding:** This study was funded through support from National Natural Science Foundation of China (Grant numbers 41525004, 41421001, 41601430) and Key Research Program of Frontier Science, Chinese Academy of Sciences (Grant number QYZDY-SSW-DQC007).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Pei, T.; Zhu, A.X.; Zhou, C.H.; Li, B.L.; Qin, C.Z. A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 153–168. [[CrossRef](#)]
2. Cheng, T.; Adepeju, M. Modifiable Temporal Unit Problem (MTUP) and its effect on space-time cluster detection. *PLoS ONE* **2014**, *9*, 1–10. [[CrossRef](#)] [[PubMed](#)]
3. Adepeju, M.; Rosser, G.; Cheng, T. Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions—A crime case study. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 2133–2154. [[CrossRef](#)]
4. Han, J.W.; Kamber, M.; Tung, A.K.H. Spatial clustering methods in data mining. In *Geographic Data Mining and Knowledge Discovery*; Miller, H.J., Han, J.W., Eds.; Taylor & Francis: London, UK, 2001; pp. 188–217.
5. Wiegand, T.; Moloney, K.A. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **2004**, *104*, 209–229. [[CrossRef](#)]
6. Lin, C.R.; Chen, M.S. Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 145–159.
7. Hinneburg, A.; Keim, D. An efficient approach to clustering in large multimedia databases with noise. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 58–65.
8. Karypis, G.; Han, E.H.; Kumar, V. Chameleon: Hierarchical clustering using dynamic modelling. *IEEE Comput.* **1999**, *32*, 68–75. [[CrossRef](#)]
9. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. 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, USA, 2–4 August 1996; pp. 226–231.
10. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. Proceedings of ACM-SIGMOD '99 International Conference on Management Data, Philadelphia, PA, USA, 31 May–3 June 1999; pp. 46–60.
11. Pei, T.; Gao, J.; Ma, T.; Zhou, C.H. Multi-scale decomposition of point process data. *Geoinformatica* **2012**, *16*, 625–652. [[CrossRef](#)]
12. Sheikholeslami, G.; Chatterjee, S.; Zhang, A. 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, USA, 24–27 August 1998; pp. 428–439.
13. Estivill-Castro, V.; Lee, I. Multi-level clustering and its visualization for exploratory spatial analysis. *Geoinformatica* **2002**, *6*, 123–152. [[CrossRef](#)]
14. Pei, T.; Zhou, C.H.; Zhu, A.X.; Li, B.L.; Qin, C.Z. Windowed nearest-neighbour method for mining spatiotemporal clusters in the presence of noise. *Int. J. Geogr. Inform. Sci.* **2010**, *24*, 925–948. [[CrossRef](#)]
15. Pei, T.; Jasra, A.; Hand, D.J.; Zhu, A.X.; Zhou, C.H. DECODE: A new method for discovering clusters of different densities in spatial data. *Data Min. Knowl. Discov.* **2009**, *18*, 337–369. [[CrossRef](#)]
16. Eberhardt, L.L. Some developments in 'distance sampling'. *Biometrics* **1967**, *23*, 207–216. [[CrossRef](#)] [[PubMed](#)]
17. Johnson, R.B.; Zimmer, W.J. A more powerful test for dispersion using distance measurements. *Ecology* **1985**, *6*, 1669–1675. [[CrossRef](#)]
18. Pascual, D.; Pla, F.; Sanchez, J.S. Non parametric local density-based clustering for multimodal overlapping distributions. Proceedings of Intelligent Data Engineering and Automated Learning—IDEAL2006, Burgos, Spain, 20–23 September 2006; pp. 671–678.
19. Prayag, V.R.; Deshmukh, S.R. Testing randomness of spatial pattern using Eberhardt's index. *Environmetrics* **2000**, *11*, 571–582. [[CrossRef](#)]
20. Schiffers, K.; Schurr, F.M.; Tielborger, K.; Urbach, C.; Moloney, K.; Jeltsch, F. Dealing with virtual aggregation—A new index for analysing heterogeneous point patterns. *Ecography* **2008**, *31*, 545–555. [[CrossRef](#)]
21. Besag, J.E.; Gleaves, J.T. On the detection of spatial pattern in plant communities. *Bull. Int. Stat. Inst.* **1973**, *45*, 153–158.
22. Ripley, B.D. Modelling spatial patterns. *J. R. Stat. Soc. B* **1977**, *39*, 172–192. [[CrossRef](#)]
23. Ashour, W.; Sunoallah, S. Multi density DBSCAN. *Lect. Notes Comput. Sci.* **2011**, *6936*, 446–453.

24. Jiang, H.; Li, J.; Yi, S.H.; Wang, X.Y.; Hu, X. A new hybrid method based on partitioning-based DBSCAN and ant clustering. *Expert Syst. Appl.* **2011**, *38*, 9373–9381. [[CrossRef](#)]
25. Pielou, E.C. The use of plant-to-neighbour distances for the detection of competition. *J. Ecol.* **1962**, *50*, 357–367. [[CrossRef](#)]
26. Getzin, S.; Dean, C.; He, F.; Trofymow, J.A.; Wiegand, K.; Wiegand, T. Spatial patterns and competition of tree species in a Douglas-fir chronosequence on Vancouver Island. *Ecography* **2006**, *29*, 671–682. [[CrossRef](#)]
27. Wakefield, E.D.; Owen, E.; Baer, J.; Carroll, M.J.; Daunt, F.; Dodd, S.G.; Green, J.A.; Guilford, T.; Mavor, R.A.; Miller, P.I.; et al. Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. *Ecol. Appl.* **2017**, *27*, 2074–2091. [[CrossRef](#)] [[PubMed](#)]
28. Flint, I.; Kong, H.; Privault, N.; Wang, P.; Niyato, D. Analysis of heterogeneous wireless networks using poisson hard-core hole process. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 7152–7167. [[CrossRef](#)]
29. Kwate, N.O.; Loh, J.M. Fast food and liquor store density, co-tenancy, and turnover: Vice store operations in Chicago, 1995–2008. *Appl. Geogr.* **2016**, *67*, 1–13. [[CrossRef](#)]
30. Neyman, J. Statistical approach to problems of cosmology. *J. R. Stat. Soc.* **1958**, *20*, 1–43. [[CrossRef](#)]
31. Spatenkova, O.; Stein, A. Identifying factors of influence in the spatial distribution of domestic fires. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 841–858. [[CrossRef](#)]
32. Teichmann, J.; Ballani, F.; van den Boogaart, K.G. Generalizations of Matérn’s hard-core point processes. *Spat. Stat.* **2013**, *3*, 33–53. [[CrossRef](#)]
33. Clark, P.J.; Evans, F.C. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **1954**, *35*, 445–453. [[CrossRef](#)]
34. Shu, H.; Pei, T.; Song, C.; Ma, T.; Du, Y.Y.; Fan, Z.D.; Guo, S.H. Quantifying the spatial heterogeneity of points. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1355–1376. [[CrossRef](#)]
35. Ripley, B.D. The second-order analysis of stationary point processes. *J. Appl. Probab.* **1976**, *13*, 255–266. [[CrossRef](#)]
36. Besag, J.E. Comments on Ripley’s paper. *J. R. Stat. Soc. B* **1977**, *39*, 193–195.
37. Kiskowski, M.A.; Hancock, J.F.; Kenworthy, A.K. On the use of Ripley’s K-function and its derivatives to analyze domain size. *Biophys. J.* **2009**, *97*, 1095–1103. [[CrossRef](#)] [[PubMed](#)]
38. Pei, T.; Wang, W.Y.; Zhang, H.C.; Ma, T.; Du, Y.Y.; Zhou, C.H. Density-based clustering for data containing two types of points. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 175–193. [[CrossRef](#)]
39. O’Haver, T. *A Pragmatic Introduction to Signal Processing*; CreateSpace Independent Publishing Platform: North Charleston, SC, USA, 1997.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).