

Article

# Identify and Delimitate Urban Hotspot Areas Using a Network-Based Spatiotemporal Field Clustering Method

Zelong Xia, Hao Li \*, Yuehong Chen and Weisheng Liao

School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China

\* Correspondence: lihao@hhu.edu.cn

Received: 16 June 2019; Accepted: 29 July 2019; Published: 31 July 2019



**Abstract:** Pick-up and drop-off events of taxi trajectory data contain rich information about residents' travel activities and road traffic. Such data have been widely applied in urban hotspot detection in recent years. However, few studies have attempted to delimitate the urban hotspot scope using taxi trajectory data. On this basis, the current study firstly introduces a network-based spatiotemporal field (NSF) clustering approach to discover and identify hotspots. Our proposed method expands the notion from spatial to space–time dimension and from Euclidean to network space by comparing with traditional spatial clustering analyses. In addition, a concentration index of hotspot areas is presented to refine the surface of centredness to delimitate the hotspot scope further. This index supports the quantitative depiction of hotspot areas by generating two standard deviation isolines. In the case study, we analyze the spatiotemporal dynamic patterns of hotspots at different days and times of day using the NSF method. Meanwhile, we also validate the effectiveness of the proposed method in identifying hotspots to evaluate the delimitating results. Experimental results reveal that the proposed approach can not only help detect detailed microscale characteristics of urban hotspots but also identify high-concentration patterns of pick-up incidents in specific places.

**Keywords:** taxi trajectory; urban hotspot; network-based spatiotemporal field; space–time dynamic patterns; concentration index

## 1. Introduction

Urban hotspots refer to regions with frequent human mobility, heavy traffic flow and prosperous economic activities. They can also reflect the characteristics and regularity of people's travel intensity in different areas [1,2]. Obtaining massive, consecutive sequences and highly accurate real-time trajectories has become increasingly possible with the increasing application of various location-based technologies [3,4]. In particular, taxi trajectory data provide a deep understanding of people's daily trip behavior and discover the mobile patterns of passengers and their underlying dynamics. Taxis' pick-up and drop-off events are the origin-and-destination (OD) pairs of taxis [5]. Such events mirror passengers' travel demands and various patterns of human mobility. Hence, pick-up and drop-off events have been widely applied to urban hotspot identification and detection. However, the typical techniques for measuring the distribution of hotspot intensity are over a 2D planar space. These approaches ignore many urban geographical phenomena associated with human activities that occur on or along the road network (e.g., points of interest, traffic crashes and street crime). In the real world, the movement in urban space is usually constrained by a dense network and the characteristic of street centrality has a large influence on the urban environment [6–8]. Thus, refining the scope of pick-up events' hotspot patterns to a 1D network space is necessary.

Spatial clustering is a major spatial data mining technique, which is used broadly for discovering hotspots in trajectory data [2,5,9,10]. In accordance with Tobler's first law of geography, closer objectives

are more related to each other. The purpose of spatial clustering analysis is to group similar objects based on their distance, connectivity and density in space [11]. This analysis commonly measures similarity among spatial objects using Euclidean distance. As mentioned previously, the pick-up and drop-off positions of passengers are normally constrained by the topology of the street network. Recent studies have focused on detecting mobile patterns of people's daily travel under the context of the urban road network. Compared with other urban data sources, GPS-enabled taxi trajectory is an important kind of spatiotemporal data. It can effectively reflect the spatiotemporal patterns of urban residents' travel behavior. Therefore, spatiotemporal clustering methods are more appropriate for urban hotspot detection. Considering the above issues, it is significant to develop a kind of spatiotemporal clustering method based on network distance to improve the accuracy of urban hotspot detection.

Existing clustering analysis approaches that use taxi trajectory data are concerned about urban hotspot detection and identification. However, studies on the delineation of urban hotspots using quantitative analysis are limited. An urban hotspot, described as the 'heart of urban activities', is usually located at the city centre and near some typical urban landmarks (e.g., main commercial street, central business district (CBD) and city square) [12]. If several high-intensity hotspots are close to each other, then there is a 'hotspot core', which can be delimited through a continuous surface. Accordingly, the statistical aggregation analysis conducted on the surface realizes the quantitative depiction of the borderline of an urban hotspot with high-density values. The present research aims to develop a method for the delineation of urban hotspots based on spatial statistical analysis.

The current study attempts to present a systemic methodology framework for identifying and delimitating urban hotspots from taxi trajectory data. Hence, our work involves four steps. Firstly, a network-based spatiotemporal field (NSF) clustering approach is proposed to identify hotspots using the pick-up and drop-off events from GPS trajectories. The NSF calculates the spatiotemporal potential value for each pick-up point. Secondly, pick-up events in an urban environment are usually constrained by street network structures. Thus the resulting values are assigned to the links based on the network segmentation algorithm. Thirdly, a continuous surface potential value can be produced by a local kriging interpolation process after obtaining the potential value of each network-based quadrat. Finally, a concentration index is presented to delimitate the urban hotspot on network-constrained centredness surface to capture the true 'hotspot core'. This index enables the clear identification of high-intensity pick-up events based on the derivation of a key isoline model. The proposed method is then tested using the taxis' pick-up events in Nanjing City, China. We explore and analyze the spatial and temporal dynamic patterns of urban hotspots based on the output results. Furthermore, we evaluate the effectiveness of our proposed NSF method in delimitating urban hotspots. The delineated hotspots based on the NSF method are more effective and reasonable than the planar-based kernel density estimation (KDE) method.

The remainder of this paper is as follows: Section 2 reviews the related work. Section 3 describes the theoretical basis of spatiotemporal data field. In Section 4 we introduce the proposed method and illustrate the detailed calculation procedure. Section 5 presents the experiment results from a case study. In Section 6 we develop an approach to delimitate urban hotspot centredness. Finally, Section 7 concludes the paper and outlines further research work.

## 2. Related Work

The latest literature has demonstrated a growing interest in applying trajectory data to detect urban hotspots. Yue et al. [9] analysed time-dependent attractive regions and mobile patterns based on pick-up data. Li et al. [13] defined hotspots as areas where the pick-up and drop-off points are clustered and discovered the spatiotemporal dynamic patterns of dwellers in these areas. Shen et al. [14] presented a grid-adaptive DBSCAN algorithm to help drivers find passengers' loading and unloading hotspots using short-dated taxi GPS traces. Inspired by the 2D Fourier transform, Pei et al. [15] proposed a density-based method for identifying two-component clusters. The clustering algorithm was applied to identify clusters of taxi trip OD data in Beijing. Zhao et al. [5] proposed

a trajectory clustering approach based on decision graphs and data fields to detect trajectory cluster centres. However, these studies mainly focus on the densely distributed areas of taxi OD points and neglect the restrictions of the street network. Hence, hotspot detection methods are currently being developed and investigated in a network space. For example, Okabe et al. [16] developed a network version of kernel density estimation (NKDE) to analyse point agglomerations in a network space and applied it to identify ‘hotspots’ of vehicle crashes. Rui et al. [17] used network K-function to investigate the spatial clustering patterns of local retail hotspots in Nanjing. Tang et al. [18] defined taxis’ pick-up and drop-off events as linear events, and presented a novel network KDE method for linear features (NKDE-L) to explore the space–time dynamics of linear features using taxi trajectory data and real street roads in Wuhan. Based on such linear representation, Zhao et al. [5] investigated a new network distance and graph partitioning-based clustering method to detect urban hotspots within the network space using GPS trajectory data, and the clustered results indicated that the proposed method can effectively identify urban hotspots. Based on the above research, urban hotspots can be defined as the cluster intensity of taxis’ pick-up and drop-off events in this study.

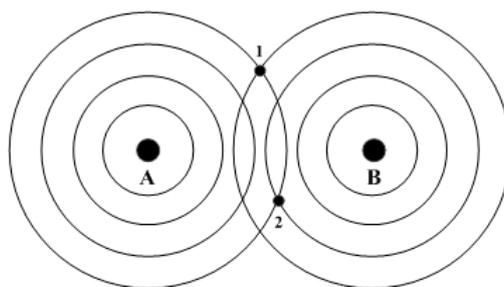
In a spatial and temporal analysis of human activities, recent advances in using taxi trajectory data to the dynamic movement detection of residents and vehicles in cities were mostly developed. Recently, Pang et al. [19] revealed a fine-grained spatial pattern based on taxi GPS data by decomposing the regularity and the disparity from point intensities. Werabhat et al. [20] constructed a spatiotemporal-varying taxi OD matrix with adaptive zoning schemes for reflecting the changing demands for taxis. While scholars have realized the importance of spatial and temporal dimensions in identifying potential ‘hotspots’ (clusters) along the network, the two are often checked separately among limited research. Therefore, network-based spatiotemporal clustering analysis can be applied to identify whether events that are close in space are also close in time. This analysis is useful in understanding the patterns and processes of various spatiotemporal point events.

### 3. Theoretical Basis

#### 3.1. Spatial Data Field

In physical space, mutual interactions among particles generate various fields. Enlightened by the field of physics, Li and Du [21] extended the field description method to abstract data objects and presented the concept of the data field. They treated an object depicted by data as a mass particle and described the mutual interaction between objects without touching each other through the aid of a data field. In the data field, each data object is regarded as a mass particle, which radiates its potential energy and is affected by others simultaneously [22,23]. When such an interaction is used, data fields can be applied to characterize the interaction among objects and mine valuable information [24]. Generally, a data field takes the following properties:

**Interactivity:** Each data object point is centred on itself and radiates outward and is then radiated by others. Figure 1 shows that data points A and B radiate data energy around themselves and interact with each other in the whole data space.



**Figure 1.** Data field property sketch-map.

**Superposition:** The potential value of each data object point is equal to the sum of energy generated at their own point in the space. Figure 1 shows that the potential value in locations 1 and 2 is the superposition sum of data objects A and B, respectively.

**Distance decay:** The potential value decreases rapidly with increasing distance, and the potential value is greater if it is closer to the field source.

Motivated by the concept of field in physics, interacting particles generate various fields. Li and Du introduced data field to describe the mutual correlation among data objects [21]. Field strength is often expressed by a potential function, which is used to calculate the potential value in arbitrary positions from data space. The specific value can quantitatively describe the spatial interactions of its neighbourhoods. In a given data space  $\Omega$ , suppose that there is a dataset  $D = \{x_1, x_2, \dots, x_n\}$ . For each data object  $x_i \in \Omega (i = 1, 2, \dots, n)$ , the potential function of data field on  $x_i$  can be calculated as [25]:

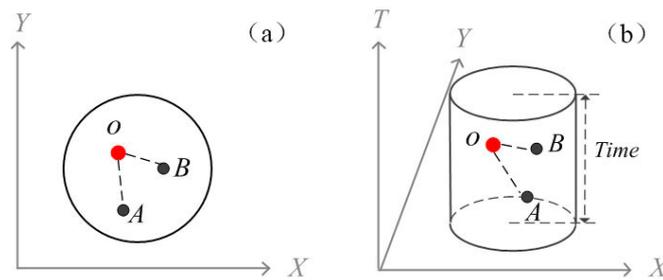
$$\varphi(x) = \sum_{j=1}^n m_j \times K\left(\frac{x_i - x_j}{\sigma}\right) \quad (1)$$

where  $x_i - x_j$  is the distance between one point  $i$  and the other point  $j$ ,  $m_j$  is the mass of data object  $j$ , the impact factor  $\sigma \in (0, +\infty)$  controls the range of interaction among objects [22],  $K(x)$  is a unit potential function to express the law of how a data object radiates its data energy in the data field, and the Gaussian kernel function is commonly employed [26].

The data field theory has been broadly applied to image classification and clustering algorithm methods. For instance, Tao et al. [27] introduced the data field method to image feature extraction. The classification of remote sensing images was conducted through steps, such as feature space construction, potential value calculation and potential image segmentation. Liu et al. [28] presented a new clustering method based on the data field in complex networks and divided network structures through the aid of the nodes' potential values.

### 3.2. Extension from Spatial to the Spatiotemporal Data Field

Space–time data refer to data that are spatial and time-varying in nature [29]. The acquisition of spatiotemporal data has become widely available due to the rapid development of positioning technology, mobile communication network and online social media. Trajectory data of taxis, especially the pick-up and drop-off points, contain substantial space and time information about the travel behavior of passengers. In the traditional data field, the distribution of field strength depends on their relative position and internal structure amongst the objects interacted with. For a series of pick-up or drop-off events, field strength is strongest when all event points in a region in space occur precisely at the same moment, and the energy decays over time. Regarding the space–time data, the intensity of the interaction between two points requires considering the time range. Figure 2 presents an illustrative example of a spatial and spatiotemporal data field. As shown in Figure 2,  $O$  is a field source centre and  $A$  and  $B$  are two referenced points and are affected by field source  $O$ . Hence, the potential value of  $O$  in space corresponds to data object mass and the interacted distance from its field source. However, space–time geographic events contain temporal behavior. The conventional potential function is none of the time dimension. For spatiotemporal data point  $O = \{x_O, y_O, t_O\}$ ,  $A = \{x_A, y_A, t_A\}$ ,  $B = \{x_B, y_B, t_B\}$  in the data space, the field strength from  $O$  to  $B$  is stronger than  $O$  to  $A$  because of the shorter spatiotemporal distance [30] (shown in Figure 2b). Thus, we must establish a spatiotemporal data field model to estimate the potential value.



**Figure 2.** Illustration of spatial field and spatiotemporal field. (a) The neighbourhood in spatial data field and (b) The neighbourhood in spatiotemporal data field.

### 3.3. Network-Based Spatiotemporal Field (NSF) Clustering Method

Taxi trajectory data are considered important spatiotemporal data sources that record the running state of taxis at some intervals. Passengers' pick-up (the status from '0' to '1') and drop-off (the status from '1' to '0') points can be extracted on the basis of the status change information. Taxis' pick-up events reflect the mobile patterns of human behavior and urban hotspots. Therefore, the data field theory can be further introduced into the trajectory data to measure taxi passengers' activity hotspots of potential distribution.

Enlightened by the previous studies on data field and spatiotemporal clustering analysis [16–18,22–25], we improve the conventional data field potential function by incorporating additional temporal weight [31].  $P = \{P_1, P_2, \dots, P_k\}$  is the taxi passengers' pick-up incident dataset, where  $P_i = \{x_i, y_i, t_i, s_i\}$ , and  $K$  is the total number of events. Each event  $P_i$  is seen as a particle with mass  $m_i$ ; in this study we assume that each activity incident has the same mass,  $P_i$  radiates its data energy and is affected by others and a virtual field surrounding it is observed. In the data space  $\Omega$ , each data point  $P_i$  is influenced by the fields from different points and is finally overlapped to obtain a superposed field. Hence, the potential value of the point  $P_i$  in the entire spatiotemporal data field  $\Omega$  can be quantitatively defined as [32]:

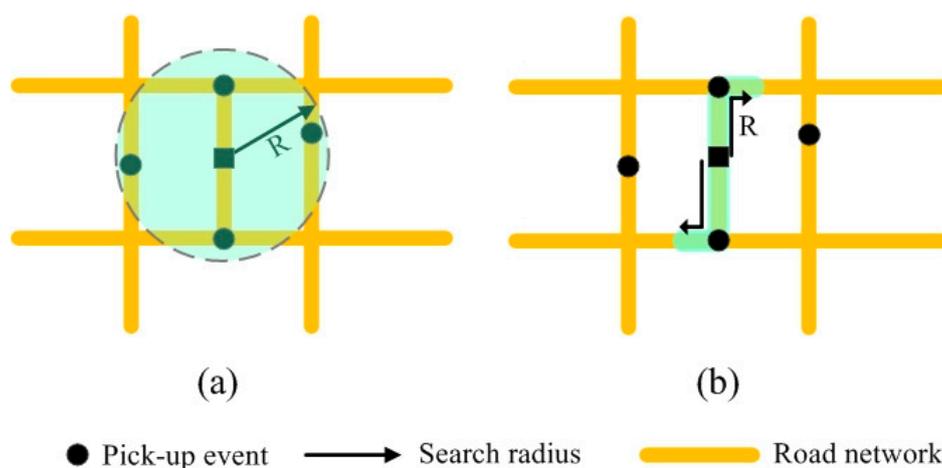
$$\varphi(P_i) = \sum_{j=1}^n \left[ m_j \times e^{-\left(\frac{d_{ij}}{R}\right)^2} \times \frac{1}{\Delta t_{ij}} \right] \quad (2)$$

where  $m_j$  represents the mass of the point  $P_j$ , the mass of each event  $P_i$  is set as 1 for convenient calculation,  $d_{ij}$  denotes the path distance between  $P_i$  and  $P_j$ , radiant radius  $R$  determines the span of distance threshold, generally it is set through the empirical experience. For instance, put the distance  $d_{ij}$  in ascending order and take the first 2% as the threshold value [25].  $1/\Delta t_{ij}$  is a normalized temporal weight coefficient and is calculated by [32]:

$$\Delta t_{ij} = \frac{\Delta t_{ij} - \Delta t_{\min}}{\Delta t_{\max} - \Delta t_{\min}} \quad (3)$$

where  $1/\Delta t_{ij}$  is the time difference between  $P_i$  and  $P_j$ , and  $\Delta t_{\min}$  and  $\Delta t_{\max}$  are for the minimum and the maximum time difference in the dataset.

While the spatial data field has been widely studied, most of them compute neighbourhood based on Euclidean distance. This process tends to overestimate the clustering tendency of network-constrained events [33,34]. Figure 3 illustrates that the search region with the Euclidean neighbourhood overestimates the number of clusters within the search radius compared with the results using network distance. Based on this concept, our work aims to improve the clustering accuracy of the traditional algorithms by using path distance.



**Figure 3.** Comparison of the search radius measured by Euclidean distance (a) and network path distance (b).

## 4. Methodology

### 4.1. Framework

In this section, we propose a systemic methodology framework and detailed process (Figure 4) for identifying and delimitating urban hotspots based on taxi trajectory data. The proposed framework includes four steps: Calculation of a spatiotemporal potential value for each pick-up point, assignment of resulting values to links, generation of regularly spaced contours in a wide range of potential values and delimitation of the centredness surfaces of the urban hotspot.

### 4.2. Calculation of Spatiotemporal Potential Value for Each Pick-Up Point

A spatiotemporal potential value for each pick-up event with the given space–time potential function is calculated (see Equation (2)). Next, pick-up points with the spatiotemporal potential attribute are achieved for expressing the intensity distribution of events.

### 4.3. Assignment of Resulting Values to Links

For taxis' pick-up events, the Euclidean distance assumption is not appropriate because the operation of taxis depends on the layout of the street network. In a 2D geographic space, geospatial data usually can be treated either as point features (e.g., traffic accidents, street robberies or service sites) or spatial units (e.g., administrative divisions, geographic boundaries or grid units) attaching some attributes (e.g., population density or car accident incidence) [35]. Similarly, network-constrained events can be considered a set of points alongside or located close to the links of the road, and their related attribute values are assigned to the links [36–38]. Moreover, the links are usually split into short segments using a network segmentation algorithm [33,38]. In this study, the methodology is used to calculate the spatiotemporal potential values of pick-up events that are expressed as line attribute-based data.

The specific implementation process is presented as follows (see Figure 5): (1) Road section segmentation. The links are divided into short segments called basic linear units (BLUs). Meanwhile, pick-up points are projected onto BLUs via nearest distance search. (2) Point feature processing. A search threshold  $R$  measured by the shortest path network distance is defined. For an arbitrary point event in a source BLU, all its neighbouring point objects within the search range  $R$  are searched (based on network distance). Then, the space–time potential value in the data space is computed with Equation (2). (3) Spatiotemporal field function calculation. For each source BLU, the space–time potential values are counted from different point events belonging to the corresponding source BLU, and finally, the resulting values are assigned to the BLUs as new attributes.

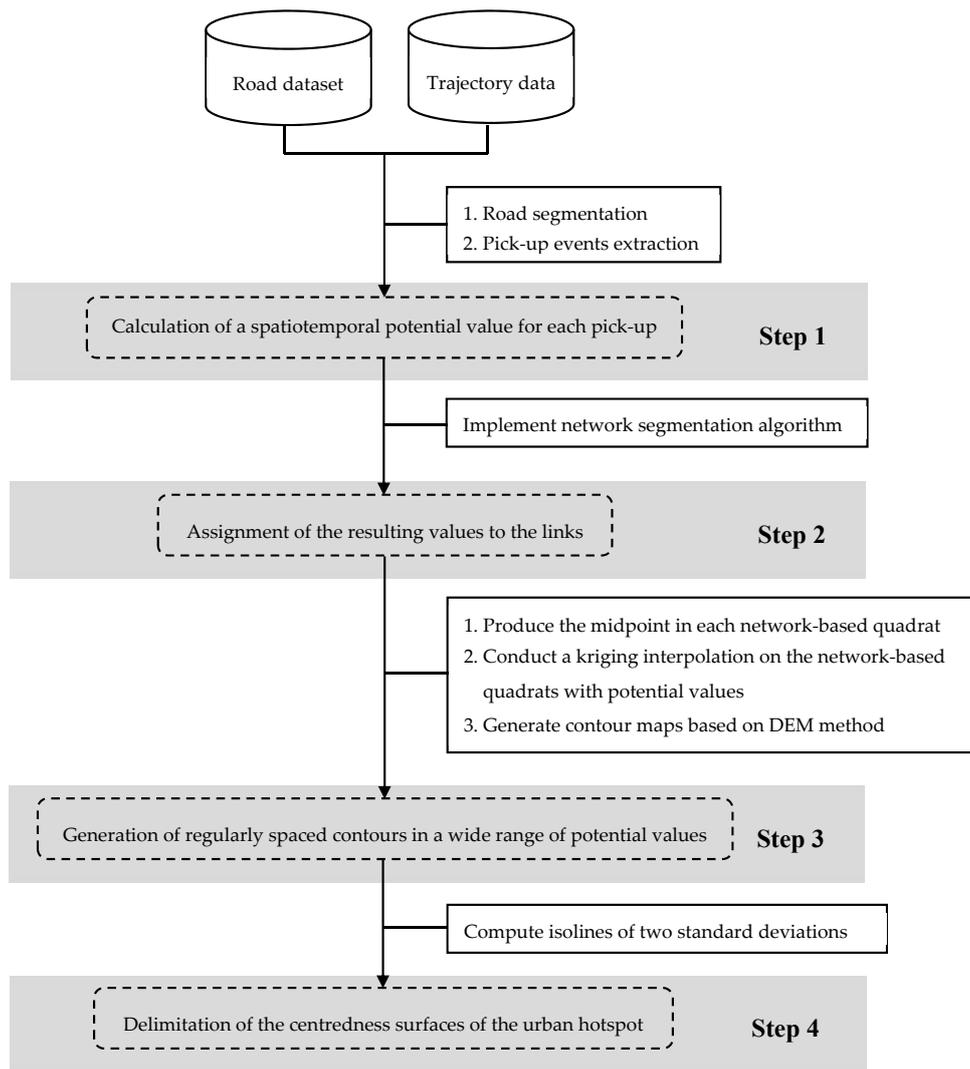


Figure 4. Frame for identifying and delimiting urban hotspot centredness.

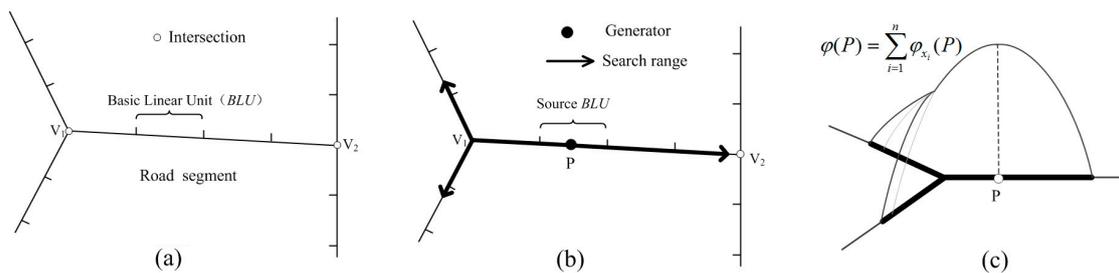


Figure 5. Algorithm implementation for spatiotemporal clustering of network events based on data field. (a) Road segments subdivision, (b) point features processing and (c) spatiotemporal field function calculation.

After calculating the potential value of each BLU, regularly spaced contours can be generated by a wide range of potential values. Firstly, the midpoint in each network-based quadrat (e.g., BLU) is produced. Secondly, a local kriging interpolation is conducted on the network-based quadrats with spatiotemporal potential values. Through such processes, a continuous surface potential value can be depicted. Thirdly, regularly spaced isolines are captured by creating a digital elevation model (DEM) based on the contour line to facilitate the manifestation of high-value characteristics.

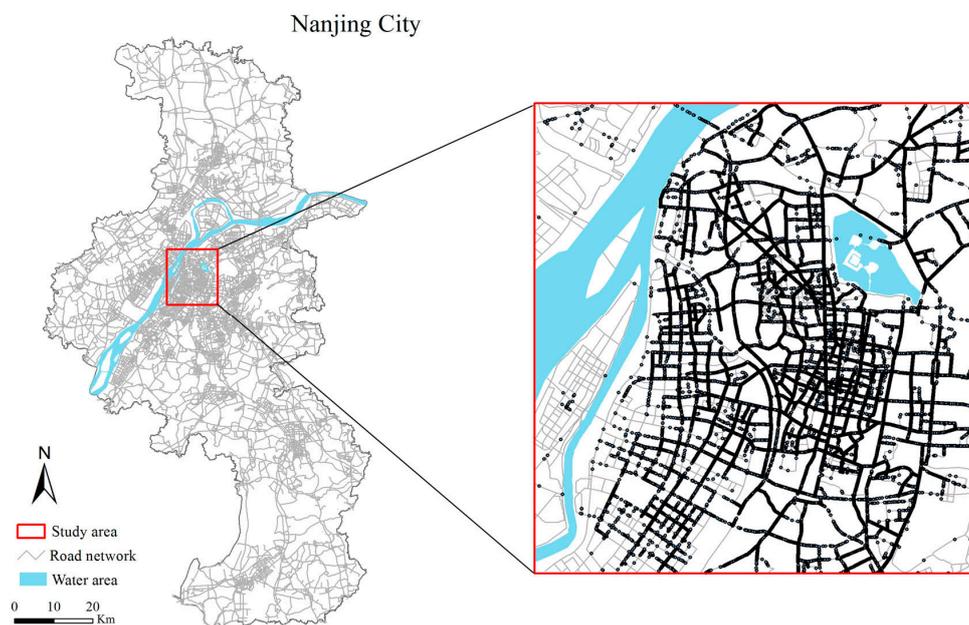
#### 4.4. Delimitation of Centredness Surfaces of Urban Hotspot

The distribution of spatiotemporal potential values based on contours is only available for capturing the fuzzy boundary of hotspots. However, delineating the true ‘hotspot core’ is difficult. A concentration index is presented to delimitate the urban hotspot on the network-constrained centredness surface. This index enables the clear identification of high-intensity pick-up events based on the derivation of a key isoline model. Specifically, we use a specified value of two standard deviation isolines to encircle hotspot regions because numerous resident travel activities are concentrated within the range.

### 5. Case Study: Exploring Spatiotemporal Clustering Pattern from Taxis’ Pick-up Events

#### 5.1. Data Description and Processing

In this work, the study area is selected from the downtown areas of Nanjing City, China, and the taxi trajectory data are acquired from local taxi companies in Nanjing. The data collection time was between the 7th and 13th of September 2015 (00:00–24:00), recording the operating of a week’s trajectories of 2927 taxicabs. Further data processing is essential to ensure data validity and achieve the goals of the main work: (1) GPS trace points are matched to road sections using the map-matching method, and (2) OD points in the trajectory are extracted in accordance with the status information (changed from ‘empty’ to ‘occupied’ or vice versa). Invalid records (noisy points or missing values) are removed from the raw data through preprocessing. A total of 2,923,198 passenger pick-up and drop-off points are extracted (as displayed in Figure 6). The road network dataset is downloaded from the OpenStreetMap website.



**Figure 6.** Study area with the pick-up and drop-off point events in Nanjing City.

#### 5.2. Experiment Settings

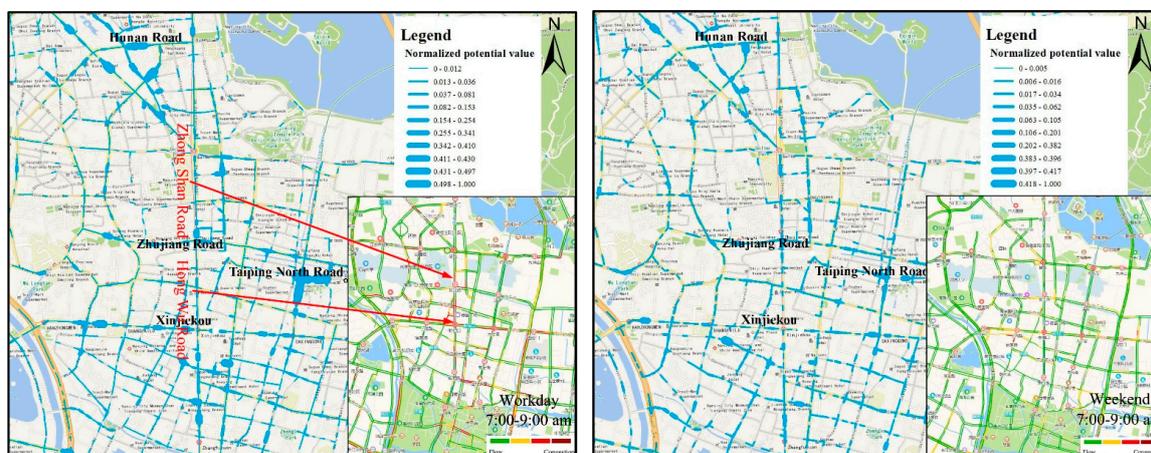
The proposed method requires the conversion of point-based pick-up events into link-based attribute data. To reveal the significant clusters of pick-up events at a relatively fine scale [39], 50 m as a basic linear unit length is enough. Similarly, neighbourhood threshold is a key parameter in structuring the pattern of network-based hotspots, which control the smoothness of the estimated output surface. An appropriate neighbourhood threshold should consider the overall distribution characteristic over the whole space and the local effects at small scales. However, few fixed rules

for threshold selection have been implemented in urban areas, and existing studies have suggested a trial-and-error process to determine the optimal one [16,36]. In view of the previous literature [39–41], scholars suggested that a 200–300 m distance is appropriate for analyzing local effects at an urban scale. Thus, a 300 m neighbourhood threshold is used in the present study.

### 5.3. Analysis of Spatiotemporal Dynamics of Urban Hotspots

Residents generally have different travel purposes on different days or times of day. Hence, passengers' pick-up events are affected by certain periods (e.g., peak or off-peak time) or different days (e.g., workdays or weekends) [2]. We investigate and compare the dynamic changes of hotspots during different periods of the day and during the same time on different days to further explore the space–time patterns of pick-up events. Two peak periods (7:00–9:00 and 16:00–18:00) and two low peak periods (20:00–22:00 and 23:00–1:00) are selected to express the temporal distribution of hotspots based on the changing regularity of dwellers' travel activities [42]. In addition, two types of days, workdays (8–9 September) and weekends (12–13 September), are applied to reflect the activity law of people on different days.

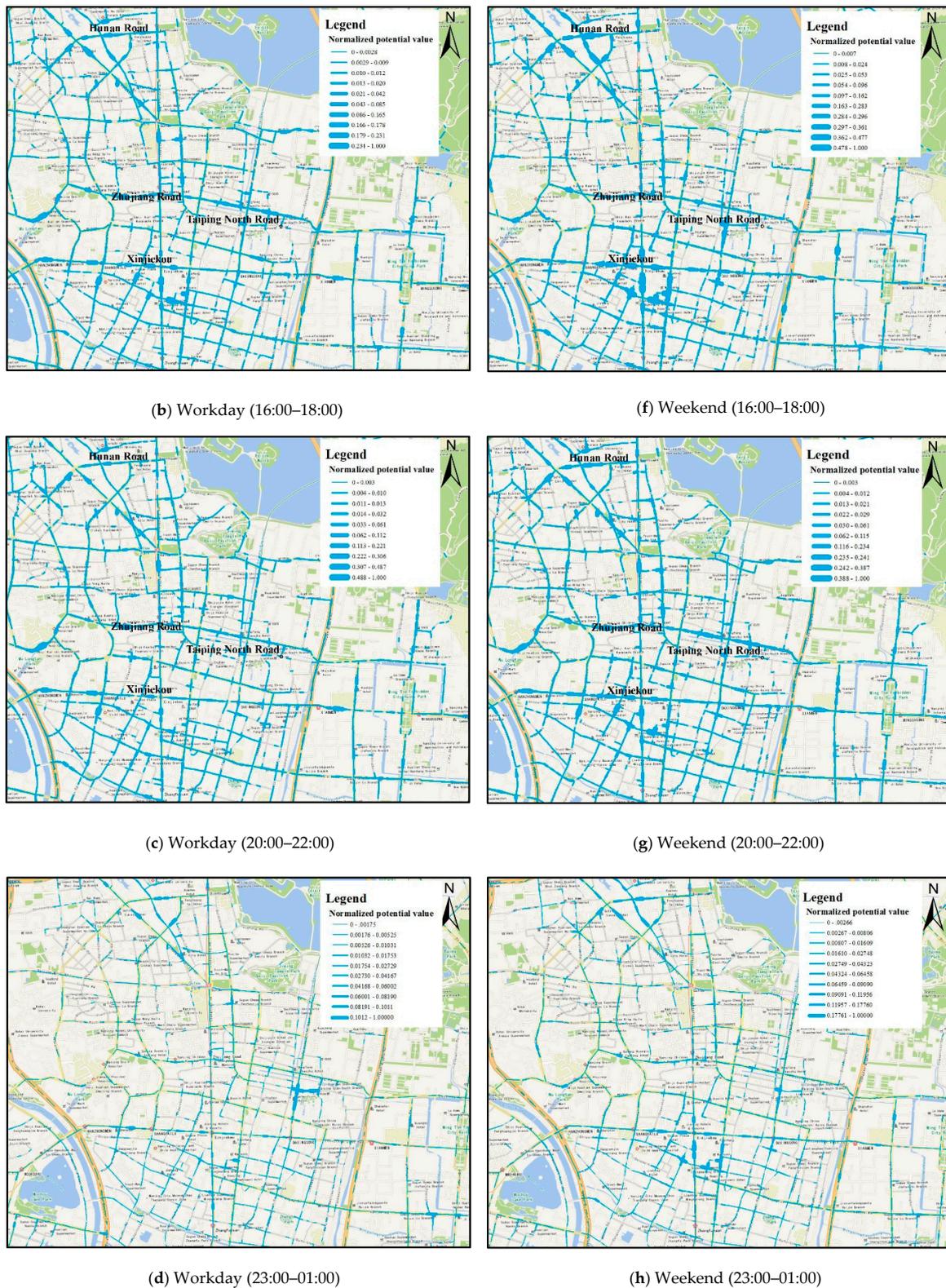
Figure 7 demonstrates that the potential value distribution of pick-up events exhibits a marked periodicity during the day whether on workdays or weekends. The potential value of pick-up events is relatively high in the daytime and slowly declines as the night comes. During 7:00–9:00 on a weekday, hotspots are mainly distributed in working places, business centres and administrative organs. For example, Taiping North Road is home to local government organs and administrative agencies, and Zhujiang Road is the biggest CBD of Nanjing with numerous enterprises. Therefore, urgent travel demand appears near these places. Some typical hotspots are mainly concentrated in commercial and leisure districts, such as Xinjiekou, Huann Road. This special distribution is closely related to shopping and leisure activities, which generally occur during the daytime and become more frequent on weekends. Furthermore, some pattern differences in the dynamics of hotspots on workdays and weekends from Figure 7 are noteworthy. During rush hours, the potential value of pick-up events in urban arterial roads (e.g., Zhongshan Road and Hongwu Road) on workdays (8–9 September) is lower than that on weekends (12–13 September). The observation is consistent with the conclusion drawn by Tang et al. [18]. Two reasons can account for this phenomenon. One is the heavy traffic conditions, which can be identified by referring to Baidu map real-time traffic information (<http://map.baidu.com/fwmap/zt/traffic/?city=nanjing>). The other factor is the subway line set along these main roads. Dwellers prefer to take the subway during this period to avoid congested traffic. These dynamic patterns also reflect the travel rules of most residents on different days and at different times of a day.



(a) Workday (7:00–9:00).

(e) Weekend (7:00–9:00)

Figure 7. Cont.



**Figure 7.** Spatiotemporal dynamic patterns of hotspots during different periods of the day and during the same time on different days: (a–d) workdays (8–9 September); (e–h) weekends (12–13 September).

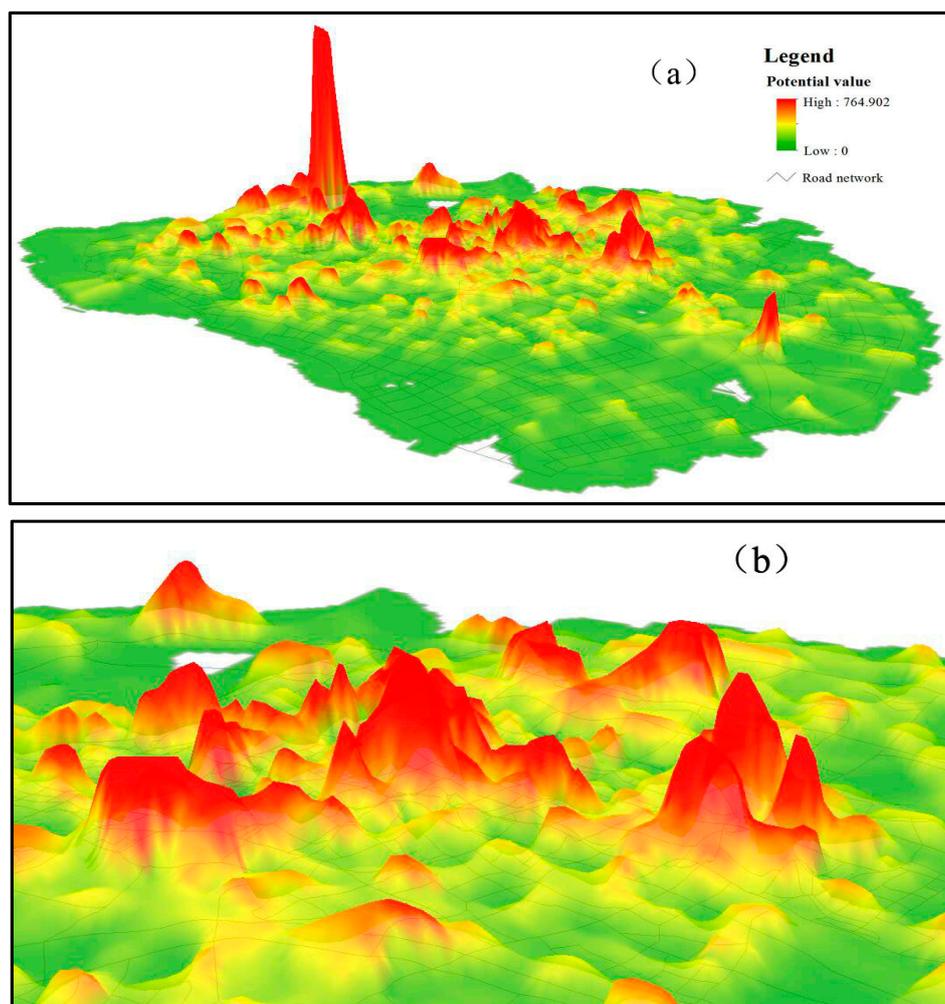
## 6. Delimitation of Urban Hotspot Centredness

In this section, a hotspot concentration index is presented to generate the isolines of hotspot centredness through 3D visualization and spatial interpolation technologies. The enclosed regions

captured by the isolines of two standard deviation values contain the highest concentration of urban activities. Hence, we use a specified value of two standard deviations to depict the centredness surface of hotspots and illustrate such effects in detail.

### 6.1. 3D Visualization of Potential Surface

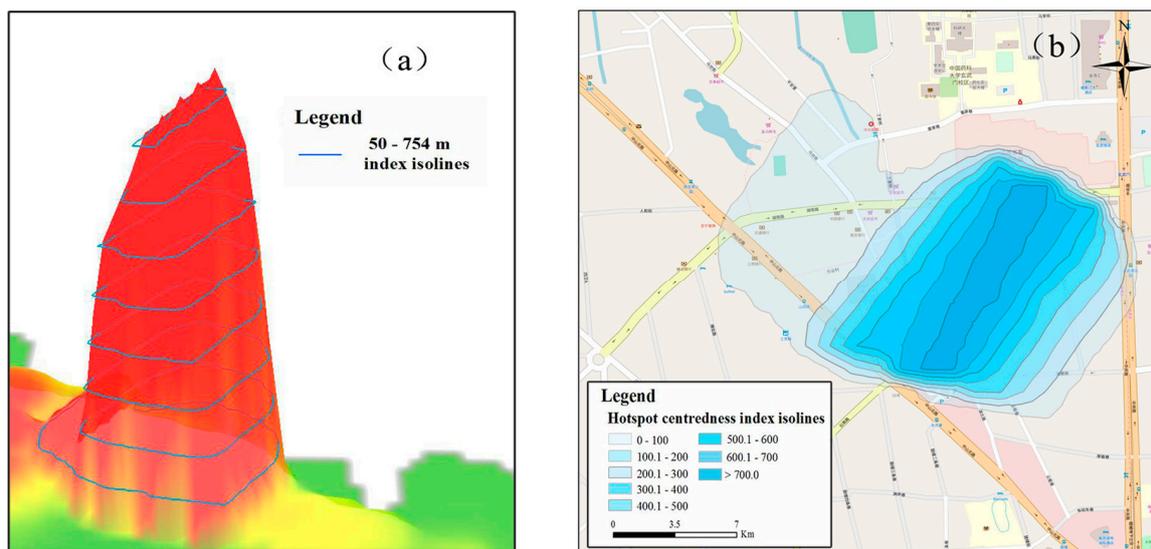
To better understand the hotspot patterns of pick-up events in a 3D space, we apply a smoothed 3D potential surface to represent the calculated hotspot intensity distribution. As mentioned above, pick-up events are abstracted as the points over a defined linear unit rather than over an area unit. Hence, in our method, the intensity value of an event is represented by an attribute of the divided linear unit. The detailed procedures for this method are as follows. The potential value of a pick-up event is firstly assigned to each BLU, and the midpoint of each BLU is generated. Then, a kriging interpolation processing is conducted on the basis of the midpoints with a BLU attribute (i.e., spatiotemporal potential value). The potential value as a height attribute is extruded into a 3D mountain shape to visualize the 2D potential surface. 'Peaks' indicate the presence of hotspots or clusters in the distribution of potential values. Figure 8 shows the 3D visualization of the potential intensity surface of hotspots in (a) the whole study area and (b) a local region. Urban hotspots manifest 'core characteristics'. Such concentration can be utilized to delineate urban hotspot shape because hotspots are usually located in the downtown area of a city.



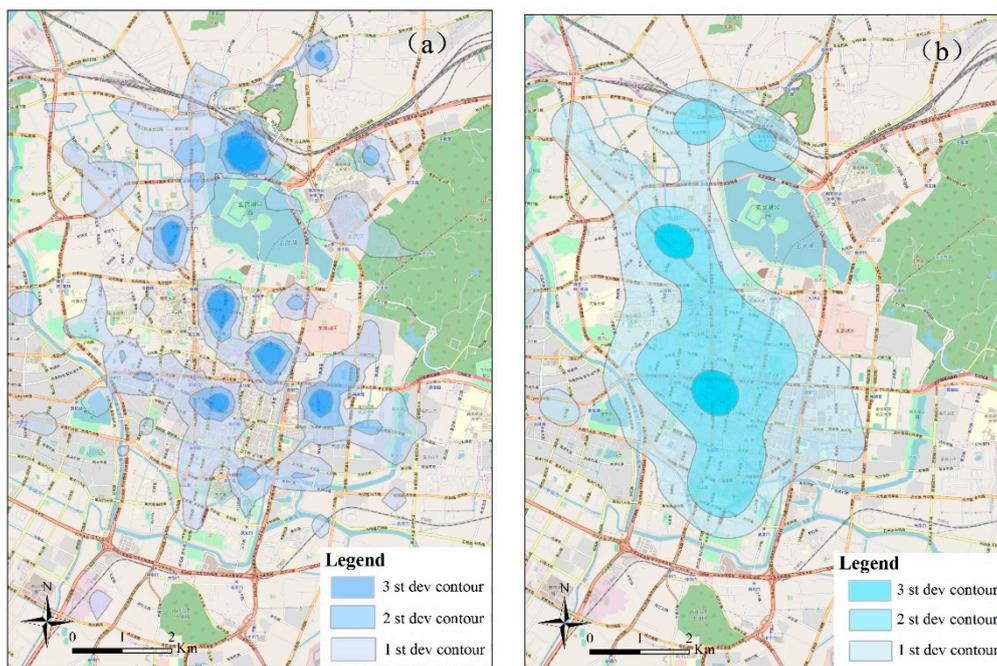
**Figure 8.** Three-dimensional visualization of the potential intensity surface of hotspots in (a) the whole study area and (b) a local region.

## 6.2. Delimitation of the Hotspot Centredness Surfaces Using an Isoline Model

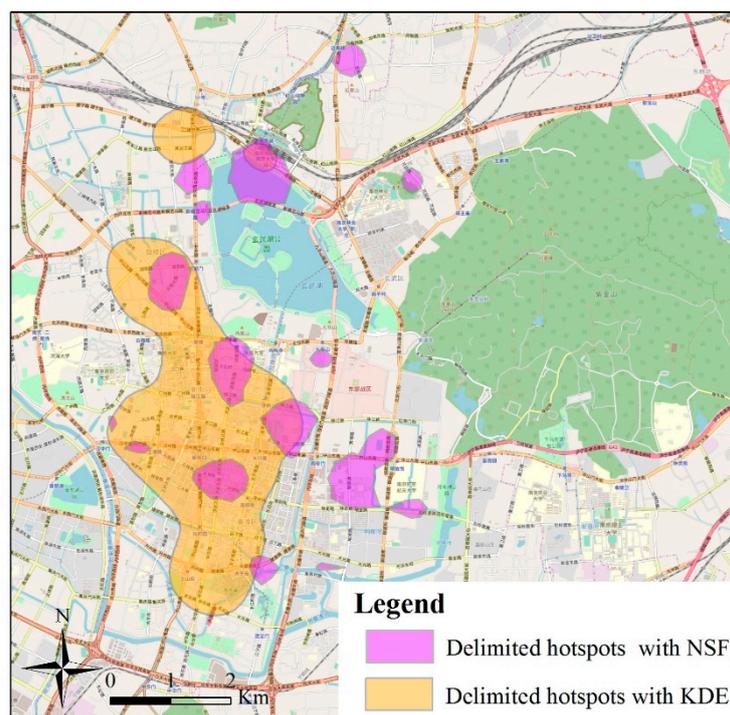
Existing studies are basically concerned with urban hotspot detection, though few pay attention to the delimitation of the urban hotspot scope. In response to the issue, in this study we propose a concentration index to delimitate hotspot centredness quantitatively. From Figure 8, the ‘peaks’ of the potential surface support the clear identification of hotspot centredness. On the basis of this consideration, contour maps can be generated from the potential attribute of each cell. Figure 9 shows the isolines of the hotspot concentration index produced by the NSF approach: (a) 3D and (b) 2D centredness surfaces. The rough scope of the urban hotspot can be captured by producing regularly spaced contours within the wide scope of potential surfaces. By producing regularly spaced contours within the wide scope of potential surfaces, the rough scope of urban hotspots can be captured. By a simple observation in Figure 9b, we can infer that the more internal contours within the enclosed area tend to have a higher concentration of traveling activities. For further examination of true hotspot ‘core’, we apply a standard deviation classification indicator to delimitate the potential surface. In view of previous research on delimitating CBD, a value of three standard deviations is commonly used [43,44]. Accordingly, in the study the standard deviation results are computed using the NSF and KDE methods to compare the delimitation differences in the network and planar spaces, respectively. As presented in Figure 10, three standard deviation isolines are generated by the network and planar methods and the output at 50 m BLU and 300 m neighbourhood threshold. Figure 10a shows the three standard deviation results of NSF. Compared with the KDE method, the aggregating results yielded from the NSF method display a network-constrained pattern (i.e., along main roads or streets). These results conform well to the distribution of pick-up events. Furthermore, the enclosed area identified by NSF can express the centredness of urban hotspots well by generating a small area according to the captured hotspot scope. By comparison, we can find that a value of a three standard deviations isolines can only capture the local ‘hotspots’ in the distribution. After several simulations, we consider that a two standard deviations isoline is most suitable for our case. Hence, we select a value of two standard deviations isolines for delimitating the hotspot scope. Figure 11 shows the shapes of the final computed urban hotspots using two different methods.



**Figure 9.** Isolines of hotspot concentration index produced by the network-based spatiotemporal field clustering (NSF) approach, results presented based on (a) 3D and (b) 2D centredness surface.



**Figure 10.** Three standard deviations isolines of urban hotspot centredness computed by two methods. (a) NSF (network-based spatiotemporal field) method; (b) KDE (kernel density estimation) method.

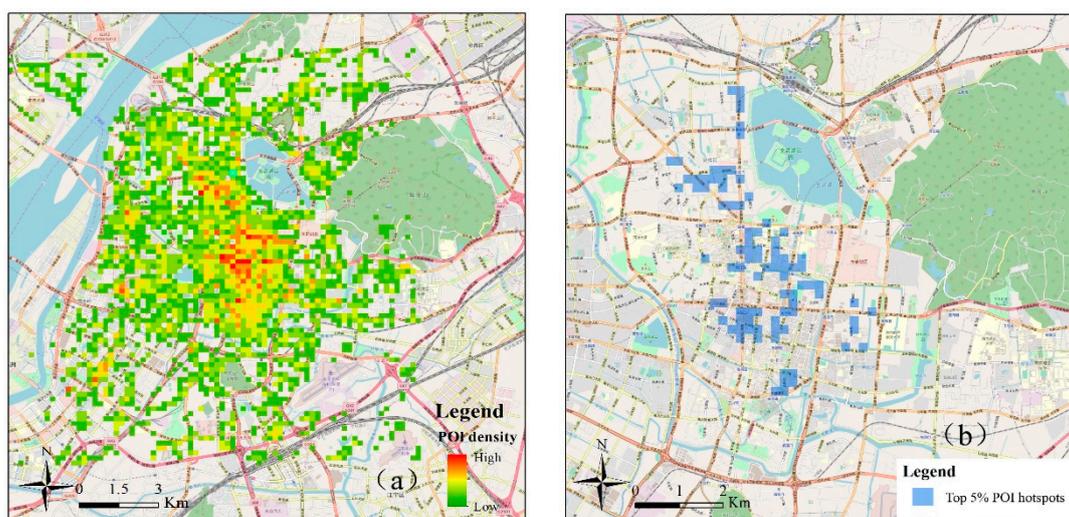


**Figure 11.** The delineated hotspots produced by the two standard deviations isolines of the NSF and KDE methods.

### 6.3. Validation of the Proposed Method

We further compare the overlapping results between delimiting scopes and the reference point of interest (POI) hotspots to investigate the effectiveness and reasonability of our proposed method. POI data include multiple categories of urban facilities, such as residential zones, restaurants, banks, entertainment venues, hospitals, hotels, stores, parks, schools and other scenic spots. These POIs reflect

the user characteristics and basic activities of urban residents (e.g., commuting, working, living and playing) [45,46]. In this manner, high-density POI areas can be understood as one of the most dominant urban centres. Therefore, we consider the high-density POI regions as the object of the comparative study. The POI data used in this work are obtained from Baidu Map API (<http://lbsyun.baidu.com/>), and all POIs are categorised in accordance with Baidu's POI standards. Data are aggregated to the appropriate spatial analysis unit in the study area to reveal the spatial characteristics of POI density. As suggested by urban geographers [47–49], a grid size of 200–300 m is suitable for a balance of fine spatial resolution in urban centres. A granular grid cell (200 m × 200 m) is selected to generate the POI density output. Figure 12a shows the distribution of POI density in the study area. Some of the identified POI hotspots can be detected by an arbitrary threshold based on the POI density results. Given the scope size of our research, we define the cells of the top 5% high-density values as identified POI hotspots (see Figure 12b).

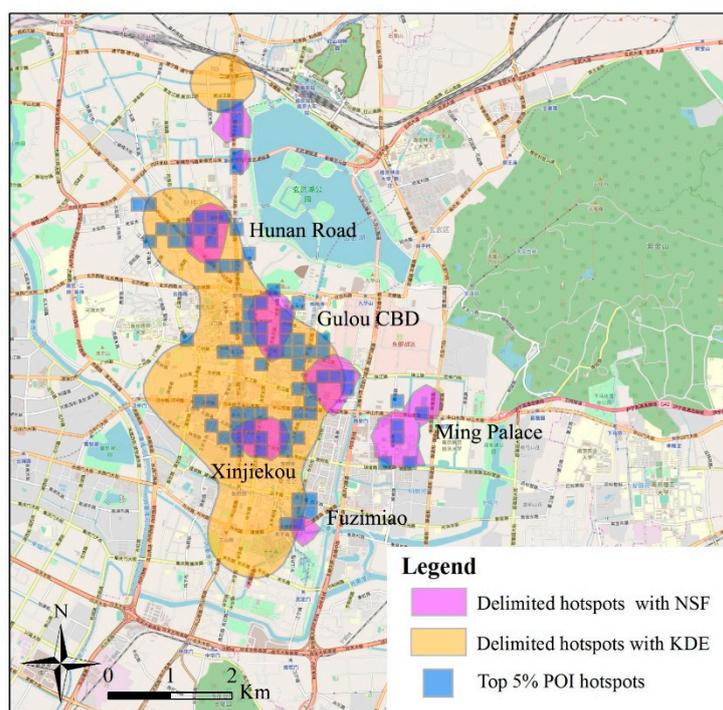


**Figure 12.** (a) Spatial distribution of POI density within urban areas of Nanjing; (b) the identified POI hotspots based on the top 5% high density value.

Figure 13 presents the delimited boundaries using two different methods and the reference POI hotspot scopes. Analysis of the cluster results in Figure 13 reveals that Nanjing's hotspots are mainly concentrated in Hunan Road, Gulou CBD, Ming Palace, Xinjiekou and Fuzimiao. These areas represent a series of urban activity centres, such as commercial streets, residential zones and scenic spots. It can be noted that the planar KDE tends to overestimate the extent of high-density pick-up events and generates a larger area for delimitation of urban hotspots compared to the network method. Based on the results, we further compare the delimitating results and evaluate the differences between the NSF and KDE methods by calculating a precision indicator as follows:

$$Precision = \frac{Area_{identical}}{Area_{delimited}} \times 100\% \quad (4)$$

where  $Area_{delimited}$  are the hotspot areas computed by the NSF and KDE methods, and  $Area_{identical}$  is the identical hotspot area between the delineated results and the reference POI hotspots. The precision index is used to calculate the ratio of the delineated hotspot areas located within the reference POI hotspots boundaries.



**Figure 13.** The overlap analysis between the delineated results and the reference POI hotspots.

Table 1 shows the evaluation results. We can find that the precision indicator for our proposed NSF (59.21%) is larger than that for KDE (32.24%). From this point, our method can refine the centredness surface of ‘hotspots’. The KDE method takes a larger space (i.e., 11.982 km<sup>2</sup>) for delimitating hotspots compared with the NSF method. However, a large hotspot area leads to small precision. In this respect, our proposed approach is more effective and accurate than the KDE method in quantifying urban hotspot centredness. By comparing the results in Figure 13 and Table 1, we can draw the following conclusions: (1) The computed urban hotspot boundaries based on our proposed NSF method can better refine the centredness surface of ‘hotspots’ compared to the area-based KDE method. (2) Considering that taxis’ pick-up points are closely associated with road networks in urban spaces, our proposed method emphasizes the constraints of road network configuration on clustering analysis. In this sense, the network-based method is more suitable for studying network-constrained urban phenomena.

**Table 1.** The evaluation of the NSF and KDE methods in terms of a precision indicator.

Statistics	NSF	KDE
Delimited hotspot area (km <sup>2</sup> )	4.097	11.982
Identical hotspot area (km <sup>2</sup> )	2.426	3.864
Precision (%)	59.21	32.24

## 7. Conclusions

Spatial cluster analysis is an important approach for identifying hotspots in various fields (e.g., transport engineering, criminology and urban planning). GPS trajectory data of pick-up and drop-off locations of taxis provide a new perspective for studying urban spatial structure and individual behavior. In addition, such data have been extensively used for detecting urban hotspots. However, the previous works in this domain are mainly conducted on a 2D plane without considering the structure of the urban road network. Hence, the present study develops NSF, which extends the notion from the spatial to the space–time dimension and from 2D planar to network spaces. This work proposes a systemic methodology framework for identifying and delimitating urban hotspots based

on taxi trajectory data. The present study first calculates a spatiotemporal potential value for each pick-up point using the NSF approach. Then, the resulting values are assigned to links. Next, regularly spaced contours are generated in a wide range of potential values, and finally, the centredness surfaces of the urban hotspot are delimited.

In the case study, the proposed method is utilized to identify city hotspots from taxis' pick-up events in Nanjing. We first analyze the spatiotemporal dynamic pattern of pick-up events during different days and times of day. For a close investigation on the centredness of hotspot distribution, a concentration index of hotspot areas is presented to refine the surface of centredness. This index supports the quantitative delimitation of hotspot areas by generating two standard deviation isolines. We then compare the overlap results between delimiting scopes and reference POI hotspot boundaries using the NSF and KDE methods, respectively. Such comparison is performed to verify the accuracy and validity of our proposed method. The precision result illustrates that the concentration ratio of NSF is larger than that of KDE in terms of the delimitation of urban hotspot centredness.

Some limitations should be considered in future research. Firstly, the experimental outputs depend strongly on the urban road network, whereas the road configuration (e.g., restricted turns, one-way roads and sidewalks) and traffic states (e.g., speed limits and road congestion) are largely simplified in this study. For example, timely transportation conditions and road attributes should be considered in a real urban environment. Secondly, the space–time distribution of urban hotspots cannot be detected completely using only taxi passengers' activity due to the complexity of urban human mobility; other urban multisensory detectors, such as web-based or smartphone apps, have not been studied yet.

**Author Contributions:** Z.X. conceived and designed the research; Y.C. and W.L. helped in data analysis and language correction; Z.X. and H.L. wrote the paper.

**Funding:** This research was funded by National Key R & D Program of China, grant number 2018YFC1508603.

**Acknowledgments:** This work was supported in part by the National Natural Science Foundation of China [No. 41830110], in part by the Fundamental Research Funds for the Central Universities [No. 2019B00314], in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province [No. KYCX18\_0616], and in part by the Fundamental Research Funds for the Central Universities [No. 2018B693X14].

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

## References

1. Zheng, Y.; Zhang, L.Z.; Xie, X. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009; pp. 791–800.
2. Zhao, P.; Qin, K.; Ye, X.; Wang, Y.A. trajectory clustering approach based on decision graph and data field for detecting hotspots. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1–27. [[CrossRef](#)]
3. Ahas, R.; Aasa, A.; Silm, S.; Tiru, M. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transp. Res. Part C Emerg. Technol.* **2010**, *18*, 45–54. [[CrossRef](#)]
4. Mao, F.; Minhe, J.I.; Liu, T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Front. Earth Sci.* **2016**, *10*, 205–221. [[CrossRef](#)]
5. Zhao, P.; Liu, X.; Shen, J.; Ming, C. A Network Distance and Graph-Partitioning-Based Clustering Method for Improving the Accuracy of Urban Hotspot Detection. *Geocarto Int.* **2017**, *3*, 1–34. [[CrossRef](#)]
6. Agryzkov, T.; Tortosa, L.; Vicent, J.F. New highlights and a new centrality measure based on the adapted pagerank algorithm for urban networks. *Appl. Math. Comput.* **2016**, *291*, 14–29. [[CrossRef](#)]
7. Agryzkov, T.; Oliver, J.L.; Tortosa, L.; Vicent, J.F. An algorithm for ranking the nodes of an urban network based on the concept of pagerank vector. *Appl. Math. Comput.* **2012**, *219*, 2186–2193. [[CrossRef](#)]
8. Agryzkov, T.; Oliver, J.L.; Tortosa, L.; Vicent, J. A new betweenness centrality measure based on an algorithm for ranking the nodes of a network. *Appl. Math. Comput.* **2014**, *244*, 467–478. [[CrossRef](#)]

9. Yue, Y.; Zhuang, Y.; Li, Q.; Mao, Q. Mining Time-dependent Attractive Areas and Movement Patterns from Taxi Trajectory Data. In Proceedings of the 2009 International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–6.
10. Chang, H.W.; Tai, Y.C.; Hsu, J.Y.J. Context-aware taxi demand hotspots prediction. *Int. J. Bus. Intell. Data Min.* **2010**, *5*, 3–18. [[CrossRef](#)]
11. Tung, A.K.H.; Hou, J.; Jiawei, H. Spatial clustering in the presence of obstacles. In Proceedings of the 17th International Conference on Data Engineering, Heidelberg, Germany, 2–6 April 2001; pp. 359–367.
12. Hollenstein, L.; Purves, R. Exploring place through user-generated content: Using Flickr tags to describe city cores. *J. Spat. Inf. Sci.* **2010**, *1*, 21–48.
13. Li, B.; Zhang, D.; Sun, L.; Chen, C.; Li, S.; Qi, G.; Yang, Q. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 21–25 March 2011; pp. 63–68.
14. Shen, Y.; Zhao, L.; Fan, J. Analysis and visualization for hot spot based route recommendation using short-dated taxi GPS traces. *Information* **2015**, *6*, 134–151. [[CrossRef](#)]
15. Pei, T.; Wang, W.; Zhang, H.; Ma, T.; Du, Y.; Zhou, C. Density-based clustering for data containing two types of points. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 175–193. [[CrossRef](#)]
16. Okabe, A.; Satoh, T.; Sugihara, K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 7–32. [[CrossRef](#)]
17. Rui, Y.; Yang, Z.; Qian, T.; Khalid, S.; Xia, N.; Wang, J. Network-constrained and category-based point pattern analysis for Suguo retail stores in Nanjing, China. *Int. J. Geogr. Inf. Sci.* **2006**, *30*, 186–199. [[CrossRef](#)]
18. Tang, L.; Kan, Z.; Zhang, X.; Sun, F.; Yang, X.; Li, Q. A network Kernel Density Estimation for linear features in space–time analysis of big trace data. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1717–1737. [[CrossRef](#)]
19. Pang, J.; Huang, J.; Yang, X.; Wang, Z.; Yu, H.; Huang, Q.; Yin, B. Discovering Fine-Grained Spatial Pattern From Taxi Trips: Where Point Process Meets Matrix Decomposition and Factorization. *IEEE Trans. Intell. Trans. Syst.* **2017**, *19*, 3208–3219. [[CrossRef](#)]
20. Werabhat, M.; Santi, P.; Merkebe, G.D.; Lina, K.; Marco, B.; Carlo, R. Constructing time-dependent origin-destination matrices with adaptive zoning scheme and measuring their similarities with taxi trajectory data. *IEEE Access* **2019**, *7*, 77723–77737.
21. Li, D.; Du, Y. *Artificial Intelligent with Uncertainty*; Chapman and Hall, CRC: Boca Raton, FL, USA, 2007.
22. Wang, S.; Gan, W.; Li, D.; Li, D. Data Field for Hierarchical Clustering. *Int. J. Data. Warehous.* **2011**, *7*, 43–63. [[CrossRef](#)]
23. Wu, T.; Qin, K. Image data field for homogeneous region based segmentation. *Comput. Electr. Eng.* **2012**, *38*, 459–470. [[CrossRef](#)]
24. Li, D.; Wang, S.; Yuan, H.; Li, D. Software and applications of spatial data mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2016**, *6*, 84–114. [[CrossRef](#)]
25. Li, C.; Ding, G.; Wang, D.; Li, Y.; Wang, S. Clustering by fast search and find of density peaks with data field. *Chin. J. Electron.* **2016**, *25*, 397–402.
26. Hug, D.; Reitzner, M. Gaussian polytopes: Variances and limit theorems. *Adv. Appl. Probab.* **2005**, *37*, 297–320. [[CrossRef](#)]
27. Jian, B.T.; Ning, S.; Zhao, Q.S. A study of the method for classification of remote sensing images based on data field cluster. *Remote Sens. Land Resour.* **2008**, *20*, 20–23. (In Chinese)
28. Liu, Y.; Jin, J.; Zhang, Y.; Xu, C. A new clustering algorithm based on data field in complex networks. *J. Supercomput.* **2008**, *67*, 20–23. [[CrossRef](#)]
29. Downs, J.A.; Horner, M.W.; Hyzer, G.; Lamb, D.; Loraamm, R. Voxel-based probabilistic space–time prisms for analysing animal movements and habitat use. *Int. J. Geograph. Inf. Sci.* **2014**, *28*, 875–890. [[CrossRef](#)]
30. Wardlaw, R.L.; Frohlich, C.; Davis, S.D. Evaluation of precursory seismic quiescence in sixteen subduction zones using single-link cluster analysis. *Pure Appl. Geophys.* **1990**, *134*, 57–78. [[CrossRef](#)]
31. Qin, K.; Zhou, Q.; Wu, T.; Xu, Y.Q. Hotspots detection from trajectory data based on spatiotemporal data field clustering. In Proceedings of the ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W7, Wuhan, China, 1–5 October 2017; pp. 1319–1325.
32. Kun, Q.I.N.; Qing, Z.H.O.U.; Yuanquan, X.U.; Wenting, X.U.; Ping, L.U.O. Spatial interaction network analysis of urban traffic hotspots. *Prog. Geogr.* **2017**, *36*, 1149–1157. (In Chinese)

33. Xie, Z.; Yan, J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. *J. Transp. Geogr.* **2013**, *31*, 64–71. [[CrossRef](#)]
34. Yamada, I.; Thill, J.C. Local Indicators of Network-Constrained Clusters in Spatial Patterns Represented by a Link Attribute. *Ann. Assoc. Am. Geogr.* **2010**, *100*, 574–594. [[CrossRef](#)]
35. Miller, H.J.; Wentz, E.A. Representation and Spatial Analysis in Geographic Information Systems. *Ann. Assoc. Am. Geogr.* **2003**, *93*, 574–594. [[CrossRef](#)]
36. Borruso, G. Network density estimation: A GIS approach for analysing point patterns in a network space. *Trans. GIS* **2008**, *12*, 377–402. [[CrossRef](#)]
37. Loo, B.P.Y.; Yao, S. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. *Comput. Environ. Urb. Syst.* **2013**, *41*, 249–261. [[CrossRef](#)]
38. Nie, K.; Du, Q.; Ren, F.; Tian, Q. A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: A case study from Wuhan, China. *Sustainability* **2015**, *7*, 2662–2677. [[CrossRef](#)]
39. Yu, W.; Ai, T. The visualization and analysis of urban facility POIs using network kernel density estimation constrained by multi-factors. *Bol. Cienc. Geodesicas* **2014**, *20*, 902–926. [[CrossRef](#)]
40. Shiode, S. Street-level spatial scan statistic and STAC for analyzing street crime concentrations. *Trans. GIS* **2011**, *15*, 365–383. [[CrossRef](#)]
41. Shiode, S. Revisiting John Snow’s map: Network-based spatial demarcation of cholera area. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 133–150. [[CrossRef](#)]
42. Yu, W.; Ai, T.; He, Y.; Shao, S. Spatial co-location pattern mining of facility Points-of-Interest improved by network neighbourhood and distance decay effects. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 280–296. [[CrossRef](#)]
43. Guo, D.; Zhu, X.; Jin, H.; Gao, P.; Andris, C. Discovering spatial patterns in origin-destination mobility data. *Trans. GIS* **2012**, *16*, 411–429. [[CrossRef](#)]
44. Borruso, G.; Porceddu, A. A tale of two cities: Density analysis of CBD on two midsize urban areas in northeastern Italy. In *Geocomputation and Urban Planning*; Murgante, B., Borruso, G., Lapucci, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 37–56.
45. Yu, W.; Ai, T.; Shao, S. The analysis and delimitation of Central Business District using network kernel density estimation. *J. Transp. Geogr.* **2015**, *45*, 32–47. [[CrossRef](#)]
46. He, Q.; He, W.; Song, Y.; Wu, J.; Yin, C.; Mou, Y. The impact of urban growth patterns on urban vitality in newly built-up areas based on an association rules analysis using geographical ‘big data’. *Land Use Policy* **2018**, *78*, 726–738. [[CrossRef](#)]
47. Jia, R.; Khadka, A.; Kim, I. Traffic crash analysis with point-of-interest spatial clustering. *Acc. Anal. Pre.* **2018**, *121*, 223–230. [[CrossRef](#)]
48. Filipe, B.E.S.; Gallego, J.; Lavalle, C. A high-resolution population grid map for Europe. *J. Maps* **2013**, *9*, 16–28.
49. Astrudmaantay, J.; Maroko, A.R.; Christopher, H. Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasyetric System (CEDS). *Am. Cartogr.* **2007**, *34*, 77–102.



© 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/>).