

Article

Grid Mapping for Spatial Pattern Analyses of Recurrent Urban Traffic Congestion Based on Taxi GPS Sensing Data

Yang Liu ¹, Xuedong Yan ^{1,*}, Yun Wang ¹, Zhuo Yang ² and Jiawei Wu ³

¹ MOE Key Laboratory for Urban Transportation Complex System Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; 12114243@bjtu.edu.cn (Y.L.); 13114248@bjtu.edu.cn (Y.W.)

² Department of Civil, Environmental, and Infrastructure Engineering, Volgenau School of Engineering, George Mason University, Fairfax, VA 22030, USA; zyang10@gmu.edu

³ Center for Advanced Transportation System Simulation, Department of Civil Environment Construction Engineering, University of Central Florida, Orlando, FL 32816, USA; wjw345178371@knights.ucf.edu

* Correspondence: xdyan@bjtu.edu.cn

Academic Editor: Bhavik Bakshi

Received: 6 March 2017; Accepted: 28 March 2017; Published: 31 March 2017

Abstract: Traffic congestion is one of the most serious problems that impact urban transportation efficiency, especially in big cities. Identifying traffic congestion locations and occurring patterns is a prerequisite for urban transportation managers in order to take proper countermeasures for mitigating traffic congestion. In this study, the historical GPS sensing data of about 12,000 taxi floating cars in Beijing were used for pattern analyses of recurrent traffic congestion based on the grid mapping method. Through the use of ArcGIS software, 2D and 3D maps of the road network congestion were generated for traffic congestion pattern visualization. The study results showed that three types of traffic congestion patterns were identified, namely: point type, stemming from insufficient capacities at the nodes of the road network; line type, caused by high traffic demand or bottleneck issues in the road segments; and region type, resulting from multiple high-demand expressways merging and connecting to each other. The study illustrated that the proposed method would be effective for discovering traffic congestion locations and patterns and helpful for decision makers to take corresponding traffic engineering countermeasures in order to relieve the urban traffic congestion issues.

Keywords: recurrent traffic congestion; traffic grid modeling; density-based spatial clustering; GPS data

1. Introduction

As contemporary GPS sensor technology enables us to track vehicle trajectories in a traffic network, it provides an alternative way to monitor traffic operation performance in a large traffic network with low cost but high efficiency. Especially, taxi floating car data (FCD) collected from installed GPS equipment presents an opportunity for the governments and scholars to detect and describe traffic congestion occurrence locations and patterns in the whole traffic network, which were previously difficult to identify due to the lack of traffic data [1].

FCD technology is a new approach for gathering traffic information, which is one of the most significant aspects in the field of Intelligent Transportation Systems (ITS). In essence, the taxi FCD data are a random sample set for the entire urban road network. There are thousands of taxis driving every day on the roads of large cities, such as New York, Boston, and Beijing [2]. More than half of the taxis have been equipped with GPS data recorders, and the majority of them work for the

whole day. The taxis' FCD technology has advantages in three aspects. Firstly, real-time traffic data can be automatically collected and sent to a processing center, which can facilitate the extraction of information about traffic conditions. Secondly, the entire road network can be treated as a collection of monitored areas. However, neither fixed sensor surveillances nor loop detectors can take charge of a large scale roadway network [3]. On the contrary, owing to the flexibility and magnitude of floating vehicles, it is possible to monitor the majority of roads in a roadway network. Finally, high-quality data can be collected with a minimum cost via GPS equipped in vehicles [4].

Previous studies have successfully applied GPS sensor devices to survey traffic for individual [5], freeway [6], and signalized intersections [7]. They are also an important data collector for location-based services (LBS) and deliver traffic data for vehicle navigation services [8]. The research and application of FCD can be summarized into three levels, namely macro, meso, and micro levels. At the macro level, FCD analysis results can be used to evaluate the effectiveness of the traffic planning implementation and assist city planners in identifying problems that were not expected during the planning stages [9]. At the meso level, the characteristics of travel speed can reflect the features of road networks [10], and the tracks of floating vehicles can even be used to discover driving route distributions and update the network structures or attributes [11]. At the micro level, FCD can be applied for urban traffic incident detection, segment demand-capacity analyses, intersection delay, and so on. For example, through monitoring traffic flow from different approaches of signalized intersections, the real-time traffic operation status at intersections can be established [12].

From the aspect of traffic management, the taxis' GPS data displayed a significant value for traffic operation performance assessment, particularly for traffic congestion identification. There are two types of traffic congestion in urban road networks: recurrent congestion and non-recurrent congestion. Non-recurrent congestion occurs owing to random events, such as traffic crashes, roadside breakdowns, or other incidents. However, recurrent congestion regularly takes place at fixed locations once traffic demand is higher than capacity [13]. If the recurrent congestions' locations, times, and intensities can be identified, the urban transportation managers can take proper countermeasures according to the congestion occurring patterns. Drivers can also benefit from the knowledge to avoid the frequently delayed routes or regions in order to reduce travel delay.

The traditional methods of traffic congestion identification are based on fixed sensors data and traffic flow analyses of the relationship among highway demands, free-flow speeds, and capacities [14]. Compared with traditional traffic detectors (e.g., loop detectors, traffic cameras, remote transportation microwave sensors), the floating car has several obvious advantages such as lower cost, wider coverage, and higher mobility. Thus, the FCD-based method has been paid considerable attention in non-recurrent congestion identification during the last decade. A temporary incident can be detected by analyzing floating vehicles' travel time of segments and acceleration noise through statistical models [15]. An abnormal traffic condition can also be reflected by the traveling speed in the road segments [16], with the average speed of floating vehicles in blocked or congested segments usually below 10 km/h [17]. By employing speed and temporal features of the segments identified on the road network, unique traffic patterns on each road can be characterized and the traffic states can be described on a segment-by-segment basis [18]. The typical algorithms that have been explored for incident or non-recurrent congestion detection include density-based clustering [19], support vector machine modeling [20], neural network-based modeling [21], etc.

FCD analyses gives substantial knowledge about traffic operation patterns of urban road networks [22]. However, few studies focused on recurrent congestion estimation using the FCD. Actually, the historical taxi FCD are especially appropriate for assessing the level of traffic congestion in the city [23] and scanning the spatial and temporal patterns of traffic congestion patterns in urban road networks because of the large scale samples in short time periods, which is helpful in being able to provide a whole and better picture of traffic situation in the traffic network. In this study, we processed taxi FCD starting with the elimination of implausible (e.g., mismatched GPS positions) and irrelevant (e.g., from taxis waiting for customers) data. Then, the trajectories floating vehicles

were used to generate road networks based on the method of grid mapping and a map matching process was conducted through matching the FCD data to the cells of the map. Further, the method of Density-Based Spatial Clustering of Applications with Noise was developed to fit traffic grid modeling analyses and cluster the congested cells. Finally, the GIS visualization technology was used for traffic congestion pattern analyses.

2. Method

2.1. Taxi GPS Trajectories Data Preprocessing

The data used in this paper was derived from the taxi-FCD system, which has been installed as the standard taxi GPS equipment in Beijing. Through the system, the data were sent to the FCD-server of the taxi company's headquarters. The time frequency of the taxi positioning collection was limited by the bandwidth of the communication channel and varied between 10 and 120 s, depending on the status of the individual taxis. In every 10 min, the data collected in each taxi were uploaded to the remote servers of the headquarters.

The data used in this study include the spatial trajectories of 12,000 taxis. It accounted for 1.3% of all motor vehicles in Beijing. Due to the characteristics of high travel frequency and long operation time, the taxi volume actually occupied up to 10 percent of the whole traffic volume in Beijing [24]. The original data were encoded by ASCII and stored in text files. The total size of the data was 15.1 GB, which was segmented into 4884 sub TXT files by 2-min time slices. The data were collected during the seven days from November 1st to November 7th, 2012.

The obtained data were saved in the format of TXT files, in which each row represents the data by recording time and the column represents the data by attributes. Table 1 lists the detailed data attributes of the FCD system. The attributes include vehicle number, taxi operation state, GPS recording time, longitude, latitude, vehicle speed, and GPS state. Especially, longitude and latitude can reflect the taxi positions, which are the basic information used to track the taxi trajectories based on the Worldwide Geodetic System 1984 (WGS84) coordinate system.

Table 1. The detailed data attributes of floating car data (FCD) system.

| Field Name | Field Type | Field Description |
|----------------------|------------|---|
| Vehicle Number | String | 6 byte characters, marking each vehicle |
| Taxi Operation State | Integer | 0 = no passenger; 1 = serving passenger |
| GPS Recording Time | Timestamp | Accurate to second |
| Longitude | Floating | Accurate to six decimal places |
| Latitude | Floating | Accurate to six decimal places |
| Vehicle Speed | Integer | Kilometer per hour |
| GPS State | Boolean | 1 = valid; 0 = invalid |

The original data's quality depends on a number of issues. The typical GPS errors may be caused by either blockage of the GPS signal or hardware/software bugs during the data collection process [25]. Previous research also found that drivers have different performances when taxis are in different operation statuses. When carrying passengers, drivers tend to drive faster and choose an optimal travel path. Inversely, drivers would drive slowly and choose a path in the higher travel-demand region to look for potential passengers [26]. Figure 1 shows that compared to the average speeds of taxis without passengers, the speeds of taxis carrying passengers were consistently higher. Therefore, the data of taxis in operation status is more applicable for the traffic analyses. Using the data, Figure 2 shows the proportion and number of taxis by the different operation statuses. The blue line represents the taxi proportion with passengers in the total number of taxis at different times. The results indicate that the proportion reached the peaks at 8:00 a.m. in the morning and 18:00 p.m. in the afternoon at workdays, but on weekends, the proportion reached a peak at 11:00 a.m. and was largely continuous. Actually, travel peaks on weekdays mainly occur during the commuting periods of the morning peak and

evening peak while the travel peaks on weekends are different from those on weekdays; on weekends, the morning peak occurs from 10:00 a.m. to 12:00 a.m. The proportion was always higher than 50%, indicating that at least 6000 taxis were carrying passengers at the same time during the whole day. It should be noted that even when the taxis are in operation status (serving passenger = 1), their behavior could be different from private passenger cars. Especially, during periods surrounding the arrival at destinations of passengers or picking up passengers at their points of origin, the taxis' speed could be continuously lower than the adjacent speed of passenger cars. In this study, if a taxi's speed is constantly close to zero in more than three consecutive observations while other vehicle's speeds are above 10 km/h, the observations would be treated as outliers and removed from analyses.

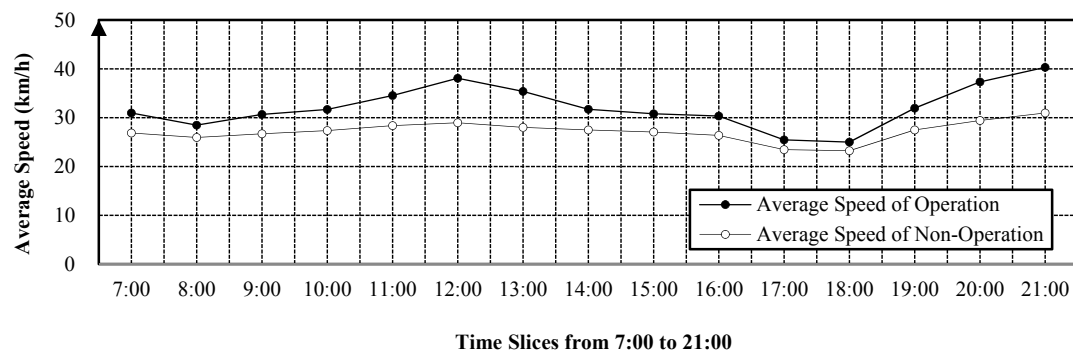


Figure 1. Average speed (km/h) of taxis of two different operating conditions.

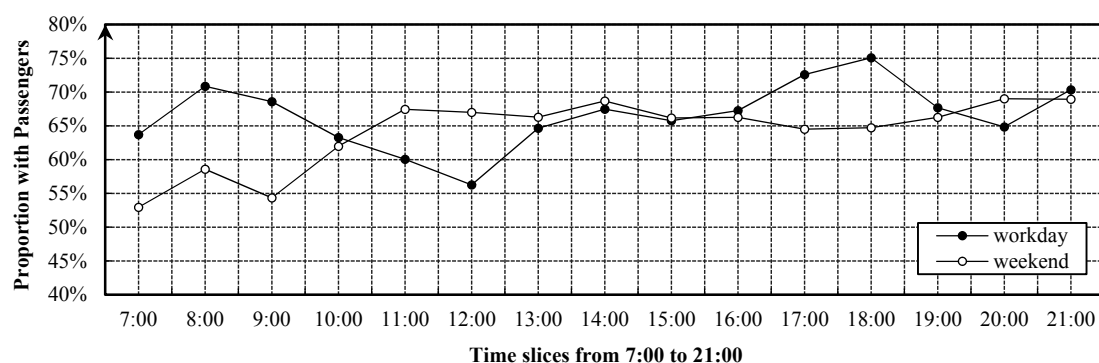


Figure 2. The number and proportion of service taxis on workdays and weekends.

The ideal situation of data acquisition frequency should be stable and concentrated. However, some errors may inevitably occur in the process of data collection and transformation due to the instability of the recording medium. Thus, it is necessary to analyze the data acquisition frequency to prepare for following analyses. The vehicle number was used as the key factor to construct a dictionary searching method. As Figure 3 shows, the taxi sampling interval of 10–15 s accounted for about 57% of the sample. The other sampling intervals were mostly more than 55 s. When the intervals were above 50 s, it indicates that the taxi service status was changing, which cannot reflect the real traffic conditions at the time. On the other hand, only 19% of the samples have a taxi sampling interval less than 10 s. Therefore, the 10–15 s sampling intervals were mostly available and reliable for the traffic analyses.

Based on the above analyses, the process of data cleansing was conducted by the following five steps: step 1 is to remove the data which contains invalid values owing to GPS errors (97% of the total data left); step 2 is to keep data that indicate taxis that are carrying passengers in operation status (39% of data left); step 3 is to eliminate the data in which the taxi speeds are out of the range from 0 to 100 km/h (34.3% of the total data left); step 4 is to remove the data in which the taxis' trajectories are beyond the traffic analysis region (we focused on the road network in the core area of Beijing, namely

the region with the fifth ring road Five Rings) (29.7% of the total data left); step 5 is to select the data with sample rates ranging 10–15 s. Finally, 25.2% of the original data was left after completing the five aforementioned steps, resulting in 48.76 million records in total.

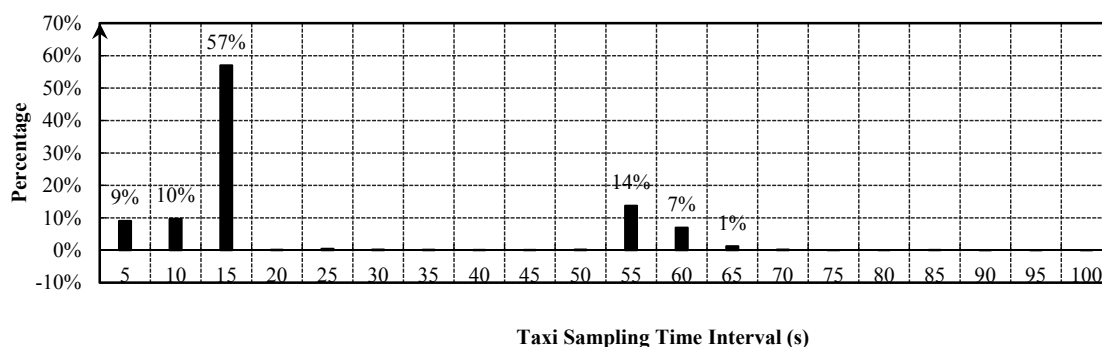


Figure 3. Distribution of the taxis' sampling time intervals.

2.2. Grid Mapping and Cell Size Identification

The traffic grid model in this research comes from the theory of city management grid modeling [27]. The city management area can be divided into grids with certain sizes either statically (the cell size is fixed for a long term) or dynamically (the cell size is varied and adjusted corresponding to different purposes of city management). The facilities within the grids will then be managed and assigned into categories so as to improve management effectiveness. Since we focused on a traffic operation performance analysis of steady road facility systems, the grid mapping and cell size should be static for uncovering recurring congestion patterns in road networks. Although linearly assigning each taxi's trajectory into road networks for traffic analyses was popular, the traffic grid models have their own advantages in two aspects. First, the sample rates are allowed to vary in a certain range and grid modeling does not require the complicated algorithms for road matching that subtly involve accuracy issues. Second, traffic grid models can be used to generalize a cell-based road network map without the need of high-quality vector GIS map.

In this study, we divided the wide-ranged urban road network into specified sizes of grids which contain massive amounts of floating car data. In our study, the road network in the core traffic area of Beijing was segmented into cells of the same size by a grid. However, choosing an appropriate cell size according to regional road properties and float car data was essential to ensure enough data for every cell while meanwhile avoiding the double count of trips. On one hand, the road attributes including expressway, arterials, sub-arterials, and collector streets in the grid should be reflected in the cell level; if the cell size was too large, the cells may contain too many road properties and the characteristics of different road networks cannot be reflected. On the other hand, if the cell size was too small, it may cause problems of offsets and leaps for continuous trajectories owing to the existing systematic uncertainties in the collection process; meanwhile, the taxi trajectory samples would be insufficient for some cells during some time periods, thus, it would make the calculations of travel speed and trajectory positions difficult.

Specifically, for the Beijing road network, if choosing a large size cell, like over 500 hundred meters, a cell would cover two parallel freeways with the same directions in a cell, which would cause the method to fail because the traffic conditions on different freeway segments cannot be distinguished. If we choose a small size cell, like tens of meters, we may not have highly dense GPS data falling into cells during a 10 min time interval, and the cell size would be smaller than the width of freeways. Additionally, the sampling rates of FCD applied in this study are 10–15 s, which means that if the cell size was chosen as tens of meters, one vehicle's speed measure points would not continuously fall into adjacent cells. Balancing the various factors, a cell size of $100 \times 100 \text{ m}^2$ was recommend for the Beijing road network's grid modeling. Accordingly, the research area was divided into a grid of

300 × 300 cells. Thus, each segment in the major road network can be expressed with continuous cells and the trajectory data can be sufficiently sampled in each cell. The geographic coordinates of the boundary for the research area and the spatial grid modeling map is shown in Figure 4. The red grid mesh represents the road segments with speed limits above 70 km/h, which are mainly expressways. In comparison, the green grid represents the road segments with speed limits under 70 km/h, which mainly contain the arterial roads, collectors, and local roads. As the network grid was developed, the grid's attribute data was structured as shown in Table 2.

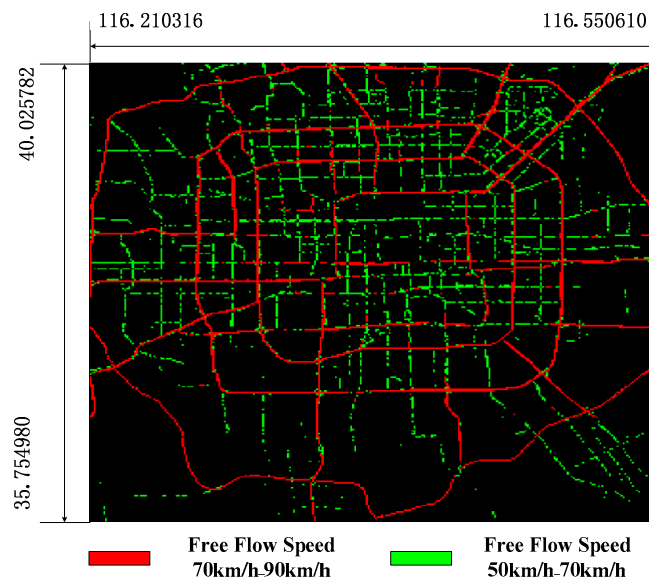


Figure 4. Spatial scale of the city in this research.

Table 2. Structure and details of the cell data.

| Field Name | Field Type | Field Description |
|------------------|---------------|--|
| Cell ID | Integer | Range of values (0–89,999) |
| Center Longitude | Integer | Accurate to six decimal places |
| Center Latitude | Integer | Accurate to six decimal places |
| Average Speed | List<Integer> | Average values of the taxi speed in cell |
| Record Number | List<Integer> | Count of the taxirecords in cell |

2.3. Cell-Based Traffic Operation Performance Index

By extracting the characteristics of traffic attributes and visually displaying them into the grid map, the traffic operation performance patterns can be figured out. The procedure includes map matching and traffic parameter estimation [8]. Map matching was to link the coordinates of vehicles with digital maps in order to locate the positions of vehicles. Traffic parameter estimation was to utilize the trajectory information, including time and speed, and identify parameters to assess the traffic congestion levels based on certain criteria, such as average speed, travel time, etc.

Firstly, the trajectory data were matched to the cells based on the GPS data, which are defined as follows:

- Cell set: $R = \{r_1, r_2, r_3, \dots, r_n\}$, $n = 90,000$
- Time slice set: $T = \{t_1, t_2, t_3, \dots, t_m\}$ in terms of every 15 min
- GPS trajectory speed data set in cell r_i during time t_j : $C^{i,j}: C^{i,j} = \{c_1^{i,j}, c_2^{i,j}, c_3^{i,j}, \dots, c_q^{i,j}\}$

Secondly, the vehicle speed performance index of cells was established for traffic operation analyses and assessment. When evaluating the road traffic congestion, the speed is an intuitive factor

reflecting traffic congestion [28]. Therefore, we define a variable $cp(r_i)$ to represent vehicle speed performance index of r_i , as shown in Equation (1), which is equal to the cell's free flow speed divided by the average operation speed for each 15 min. As the operation speeds of cells can be easily obtained according to the GPS data, the free flow speed $r_i(\bar{\mu}_F)$ is identified as the 95th percentile of operation speed distribution.

$$cp(r_i) = \sum_{j=1}^m \frac{r_i(\bar{\mu}_F)}{c^{i,j}}, \quad (1)$$

In order to intuitively demonstrate the congestion degree of the road network, the Min-Max normalized method was used to transfer the congestion indexes into the range between 0 and 100, in which the value of 100 indicates the most congested condition and the value close to 0 indicates free flow condition.

2.4. Density-Based Spatial Clustering Algorithm for Traffic Congestion Pattern Analyses

In this study, the method of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was applied to cluster the congested cells, and the GIS visualization technology was used for traffic congestion pattern analyses. DBSCAN can help to find a number of clusters starting from the estimated density distribution of corresponding nodes [29]. It groups the cells that are closely reachable as well as separates the groups with different attributes. The DBSCAN method can be well applied to data without prior knowledge about the number of classifications. The key definitions of DBSCAN are shown as follows:

- ϵ neighborhood: a circle with a radius of ϵ around a data point, which is also called the threshold between data points.
- minPts: the minimum number of data points needed within a ϵ neighborhood.
- Core points: data points are defined as core points when the number of valid data points within the ϵ neighborhood is larger than the value of minPts.
- Border points: data points are defined as border points when they are reachable by certain core points but their number within the ϵ neighborhood is smaller than the value of minPts.
- Outlier: it is neither a core point nor a border point.

In this study, we proposed DBSCAN algorithms for fitting traffic grid modeling analyses. As cells are analyzed in the grids-based network, the data points in DBSCAN are redefined as cells. First, the ϵ neighborhood is defined as the number of adjacent cells per direction around the core cell, which is equivalent to the radius of the traffic analysis zone. Second, we used the sum of speed performance index to quantify minPts within a ϵ neighborhood, as shown in Equation (2), where for a cell r_i , all the neighbor cells that are within the radius ϵ are composed of the set of $\{r_{i1}, r_{i2}, r_{i3}, \dots, r_{ip}\}$.

$$SCI(r_i) = \sum_{l=1}^p cp(r_{il}), \quad (2)$$

As illustrated in Figure 5, minPts is defined as the threshold value of $SCI(r_i)$. Then, the Core Cells are defined as the cells when the $SCI(r_i)$ is higher than minPts; the Border Cells are defined as the cells when the $SCI(r_i)$ is lower than minPts but reachable by the other Core Cells; and the Outlier Cells are those unreachable by any other cells.

The process of the algorithm is shown in Figure 6. After the road network is segmented into grid cells, the $SCI(r_i)$ is assigned into each cell, and iterative procedures are needed to categorize the cells into different clusters $Class_ID(r_i)$.

- Step 1: $Class_ID(r_i)$ is initialized as the value of -1 and the attribute of each cell is coded as false, $isVisited(r_i) = \text{false}$.
- Step 2: Traversing the whole grid, if the $SCI(r_i) > minPts$, the cell r_i is identified as a core cell;

- Step 3: If r_i is a core cell and $isVisited(r_i) = \text{false}$, traversing core cells within the ε , set $isVisited(r_i)$ for all of the cells within the ε neighborhood to be true;
- Step 4: Moving to the next core cell within the current ε neighborhood and updating its neighborhood. Repeating Step 3 and traversing all of the potential core cells until $isVisited(r_i)$ for all of the core cells is true. Assigning $Class_ID(r_i)$ for all the visited cells to be $Class_ID$.
- Step 5: Identifying any unvisited core cell, repeating Step 3 and Step 4, updating $Class_ID = Class_ID + 1$ until $isVisited(r_i)$ for all of the core cells through the grid is true.

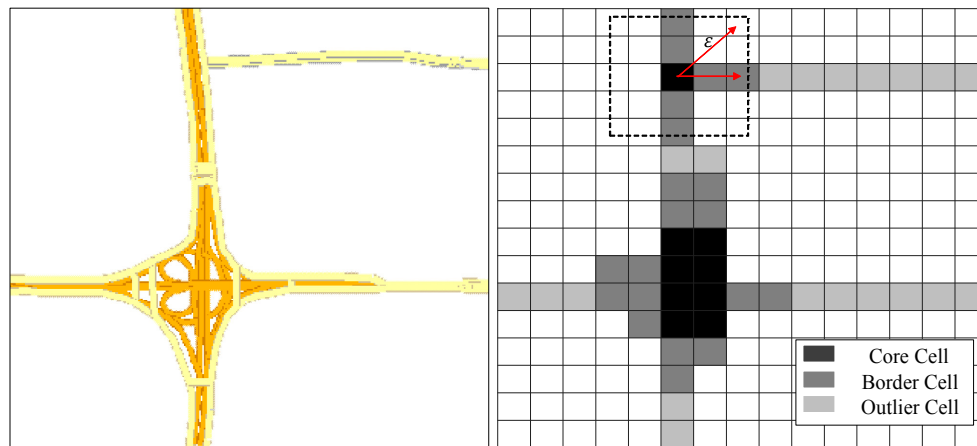


Figure 5. Schematics of the variable definition.

| | |
|--|---|
| DBSCAN of grid-based congestion area estimation: | |
| Input: SCI of all cells, radius ε , $minPts$; | |
| Output: clusters that meet the requirement. | |
| 1. | Initialize the $Class_ID(r_i)$ of cell r_i to be -1, indicating not belonging to any of the clusters. Set the visited history parameters ($isVisited(r_i)$) to be false. |
| 2. | Initialize radius ε and $minPts$, and set n as the number of data points. |
| 3. | For i from 1 to n |
| 4. | If $SCI(r_i) > minPts$ |
| 5. | Mark cell r_i as core cell |
| 6. | End if |
| 7. | End for |
| 8. | For i from 1 to n |
| 9. | $Class_ID = Class_ID + 1$ |
| 10. | If $isVisited(r_i) = \text{false}$ && cell r_i is a core cell |
| 11. | $Class_ID(r_i) = Class_ID$ |
| 12. | set $isVisited(r_i) = \text{true}$ |
| 13. | While there is unvisited cell r_j within the ε neighborhood of r_i |
| 14. | set $isVisited(r_j) = \text{true}$ |
| 15. | set $Class_ID(r_j) = Class_ID(r_i)$ |
| 16. | If cell r_j is core cell |
| 17. | enumerate other unvisited cells in ε neighborhood |
| 18. | End if |
| 19. | End while |
| 20. | End if |
| 21. | End for |

Figure 6. Pseudo-code of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm for clustering cells.

3. Modeling Results and Discussions

3.1. Visualization of Traffic Operation Performance Index

In this study, the ArcGIS software was applied to formulate both 2D and 3D maps of the road network by assigning the traffic operation performance indexes into the cells, referring to the previous research [13]. Figure 7 displays the traffic operation performance distribution patterns during 2 h, respectively for the time periods of 7:00–9:00 a.m., 9:00–11:00 a.m., 14:00–16:00 p.m., 17:00–19:00 p.m., 19:00–21:00 p.m., and 22:00–24:00 p.m. We can visually tell that the morning peak's traffic distribution in the network (7:00–9:00 a.m.) is obviously different from the afternoon peak (17:00–19:00 p.m.), during at which point the expressway system is more congested, especially in the west and east road segments of the second ring road and the east segment of the third ring road. During the daytime non-peak hours (9:00–11:00 a.m. and 14:00–16:00 p.m.), the traffic operation performance's spatial patterns are similar and traffic congestion is scattered in the networks within the fourth ring road. During the nighttime, the traffic conditions gradually become smooth from the period of 19:00–21:00 p.m. to the period of 22:00–24:00 p.m. Further, Figure 8 displays a 3D image of Beijing's expressway network's average traffic operation performance distribution map over the whole day. The height and color in the grid map represent the different congestion levels of cells, which can be easily utilized to find out the severe congestion areas. Particularly, the higher and darker a cell is, the higher the cell's congestion level will be. It shows that the closer the road segments are located in the center of the expressway network, the higher the traffic operation indexes would be. Although the analyses were focused on using the historical FCD to assess traffic operation performance, the technique applied in this study can be transferred to the purpose of dynamically monitoring the network's traffic conditions if the taxi FCD are input into the system in a real-time manner.

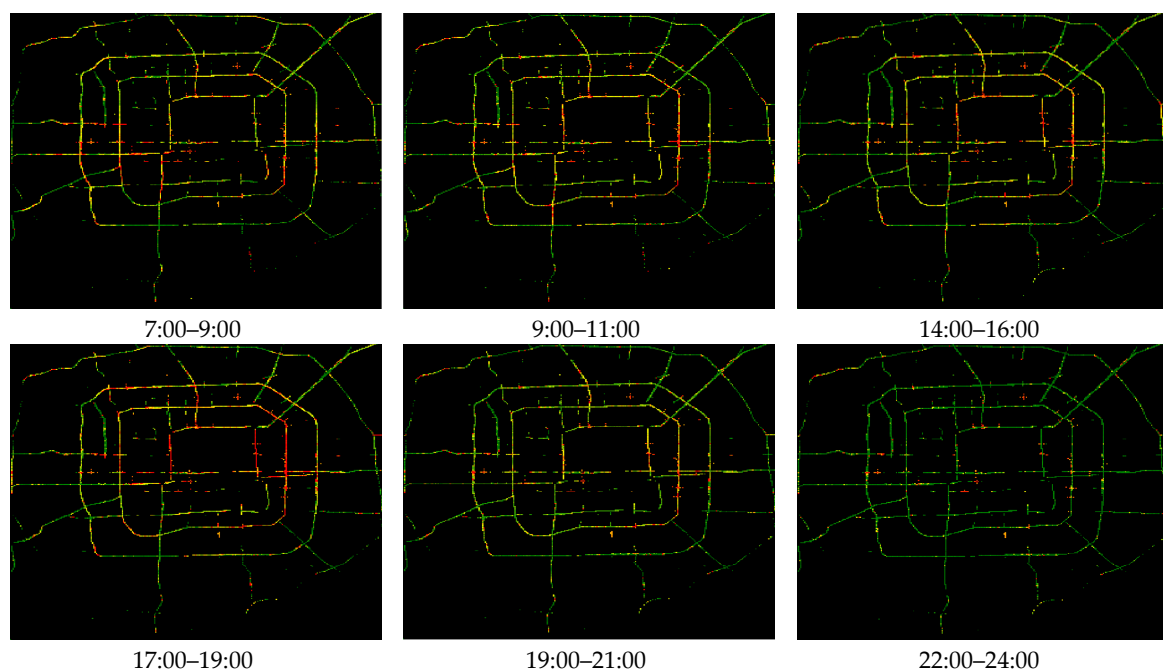


Figure 7. Traffic operation performance distribution patterns during 2 h.

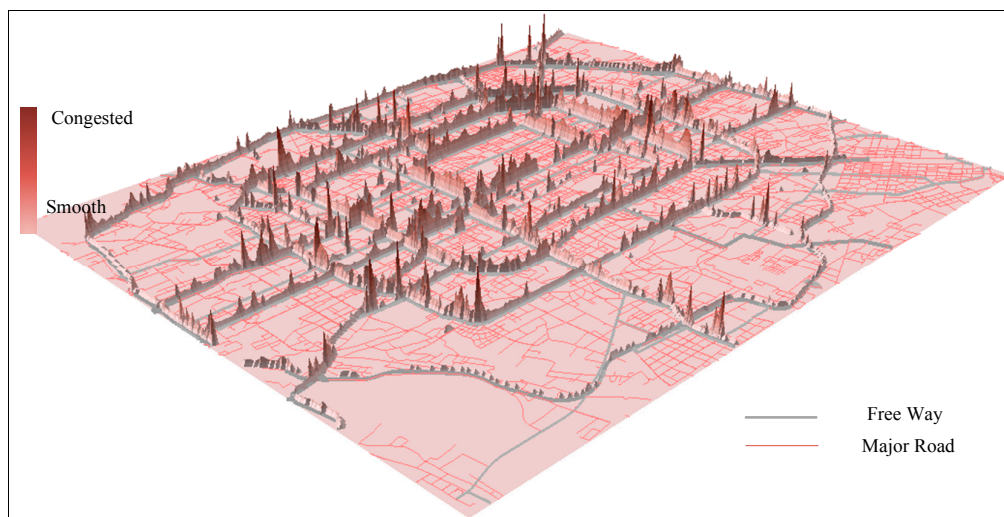


Figure 8. 3D Visualization of Traffic Operation Performance Index in Beijing.

3.2. Parameter Analyses of ϵ . and $minPts$.

After obtaining the traffic operation performance indexes for all cells, the DBSCAN algorithm was applied to detect traffic congestion patterns. In DBSCAN, there are two vital parameters, i.e., $minPts$ and ϵ , to evaluate whether cells should be clustered. A discussion about how to set these two variables is elaborated as follows.

Setting the value of ϵ determines the traffic attributes' homogeneity of the clusters. If the ϵ is too small, such as one cell neighborhood (which means a 100 meters radius around a core cell), it would cause the cells with very close traffic characteristics to be classified into different clusters. For example, the cells representing the intersection approaches for different directions could be separated into several small-scale discrete clusters, rather than one cluster. On the other hand, if the value of ϵ is too large, the adjacent parallel road segments in the road network would be classified into one cluster, resulting in many irrelevant cells falling into one giant traffic congestion region. For example, when ϵ was set as an eight cell neighborhood, the congestion region scope would be as large as 10 km, as shown in Figure 9. Eventually, it was found that when ϵ was set as a two cell neighborhood, the clusters could both reflect the traffic features and identify the detailed traffic congestion occurrence patterns.

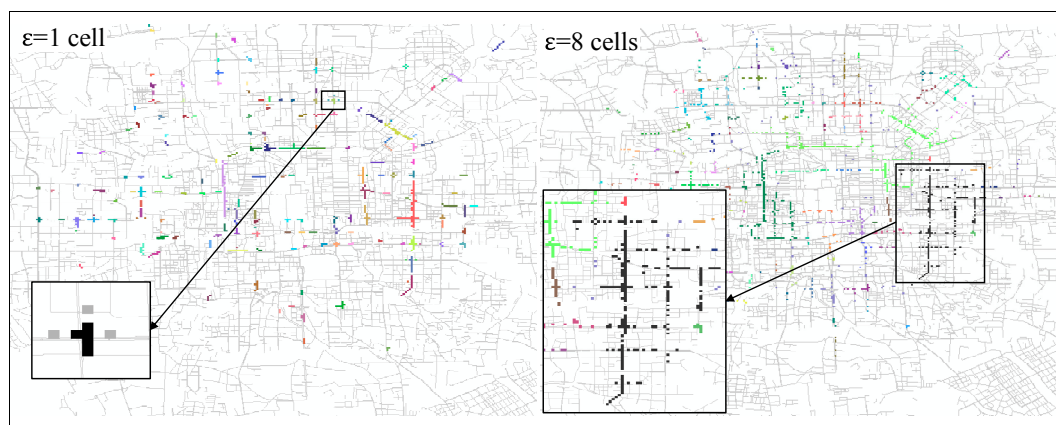


Figure 9. Congestion region scope with different value of ϵ .

A small value of $minPts$ can cause more adjacent cells being clustered into one class. Meanwhile, the total number of classes will be less. With the value of $minPts$ increasing, some original classes will

be split into more classes and the total number of classes will increase. However, when the value of $minPts$ continues to increase, more cells will be identified as outliers, resulting in that only the cells with high SCI can be kept into the clusters and the total number of classes will decrease.

To select the appropriate value of $minPts$, the corresponding clustering number based on each existing value of $minPts$ was calculated as shown in Figure 10. The results show that one class will be generated when the value of $minPts$ is more than 370. Then, as the value of $minPts$ increases, the number of classes increases. When the values of $minPts$ are within the range of 890–1050, the number of classes becomes stable in the range of 50–55. When the value of $minPts$ is 1050, the number of classes starts to drop down. Finally, no classes will be generated once the value is more than 1550. Optimally, when the value of $minPts$ is equal to 960, the number of generated classes reaches the maximum value, namely 55 classes. In this study, we chose the maximum number of clusters in order to identify different types of traffic congestion patterns presenting in the road network.

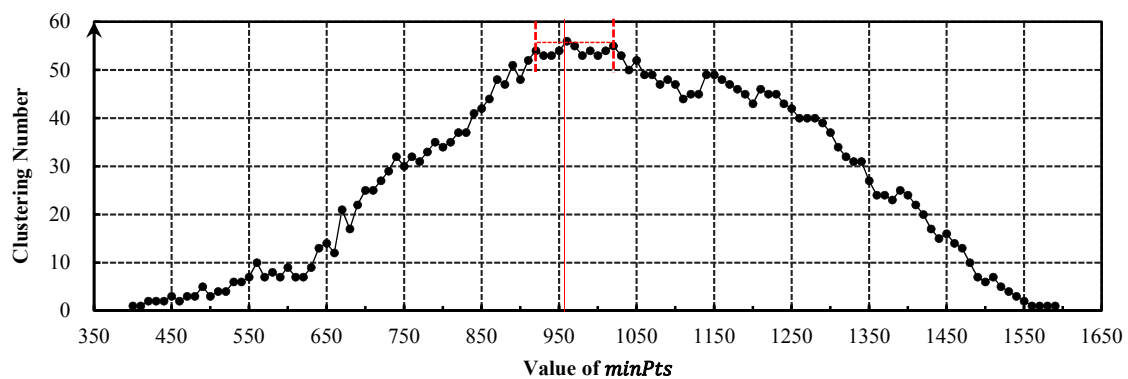


Figure 10. Clustering classes and value of $minPts$.

3.3. Traffic Congestion Pattern Analyses

When setting the value of ϵ as 2 cell neighborhood and the SCI value of $minPts$ as 960, Figure 11 shows the spatial distribution of the 55 traffic congestion clusters in the grid map. It indicates that the traffic congestion patterns can be categorized into three types, namely point type, line type, and region type, as summarized in Table 3.

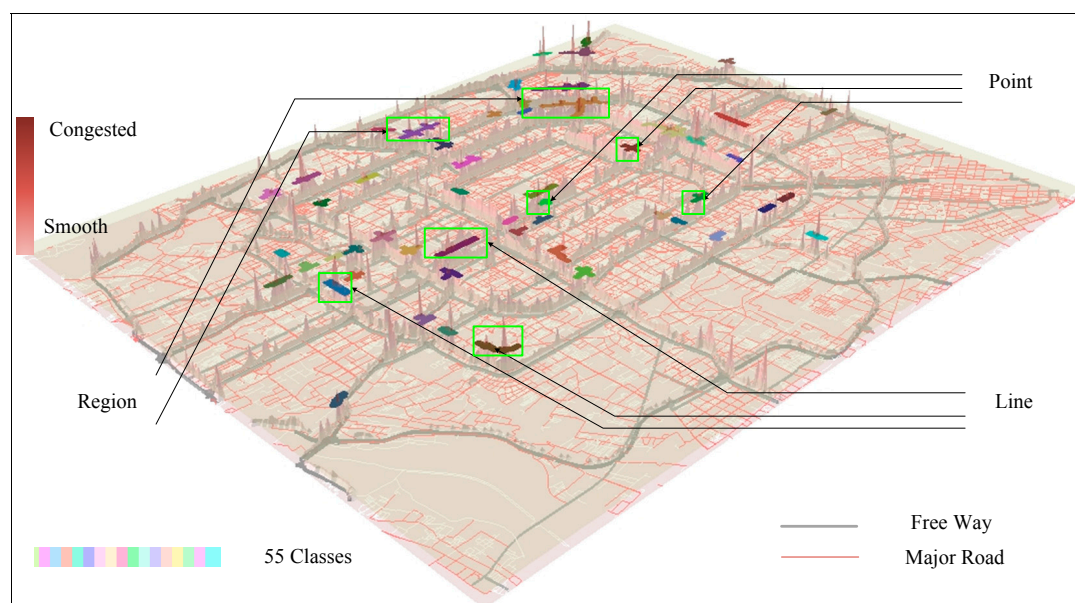
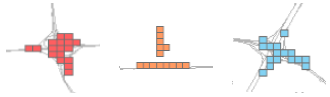

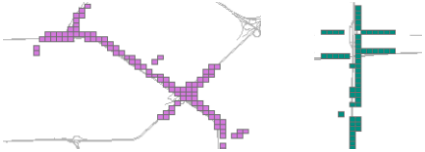


Figure 11. 3D visualization of congested cells clustering result.

Table 3. Spatial patterns of point, line, and region congestion.

| Spatial Pattern | Critical Causation | Typical Examples |
|-----------------|---|--|
| Point | Insufficient capacities at the nodes of the road network. |  |
| Line | High traffic demand or bottleneck issues in the road segments. |  |
| Region | Multiple high-demand expressways merge and connect to each other. |  |

Point-type congestion occurs at intersections of arterials or interchanges of expressways. This kind of congestion is usually caused by insufficient capacities at the nodes of the road network. It is independent and isolated from the other congested areas. Improving geometric design for interchanges and optimizing intersection signal timing design can both represent effective countermeasures to relieve the point-type congestion [30]. A serious point congestion may also lead to linear congestion, which typically emerges in the urban expressways or arterials of the segments. Line congestion is often caused by high traffic demand or bottleneck issues. In order to relieve linear congestion, an advanced traffic guidance system that delivers traffic information to drivers in a real-time manner can be applied for balancing the traffic demand distribution in the local road networks [31]. Regional congestion usually occurs and covers the whole local road networks when multiple high-demand expressways merge and connect to each other. In the Beijing road network, the most typical regional congestion exists in the areas where the third ring road and fourth ring road intersect with the high speed expressway to the National Airport. Guiding the traffic from the airport expressway to the other alternative routes or encouraging travelers to or from the airport to use mass transit systems would be proper solutions in resolving the regional congestion issues [28].

Furthermore, Figure 12 shows the distribution of the number of cells in the 55 clusters. Most of the clusters contain 10–14 cells, representing 1–1.4 km congestion mileage; some heavy congestion clusters contain 26–32 cells, representing up to 2.6–3.2 km congestion mileage, which frequently occurs in the road network. As shown in Figure 13, a strong correlation is found between the variable *SCI* and the number of cells. Intuitively, the total value of *SCI* within a cluster increases with an increase in the number of cells. However, Figure 14 illustrates the values of total *SCI* in most clusters are less than 18,000, while only six clusters' total *SCI* values are larger than 18,000 and their average *SCI* values are larger than 900 at the same time. Interestingly, using the total *SCI* value 18,000 and average *SCI* value 900 as the division line, the 55 clusters can be separated into three groups, which just match then explain the spatial patterns of point, linear, and regional congestion. The results indicate that the point type clusters have a more intensive degree of congestion but a lower influencing scope; however, the regional type clusters have both an intensive congestion degree and scope.

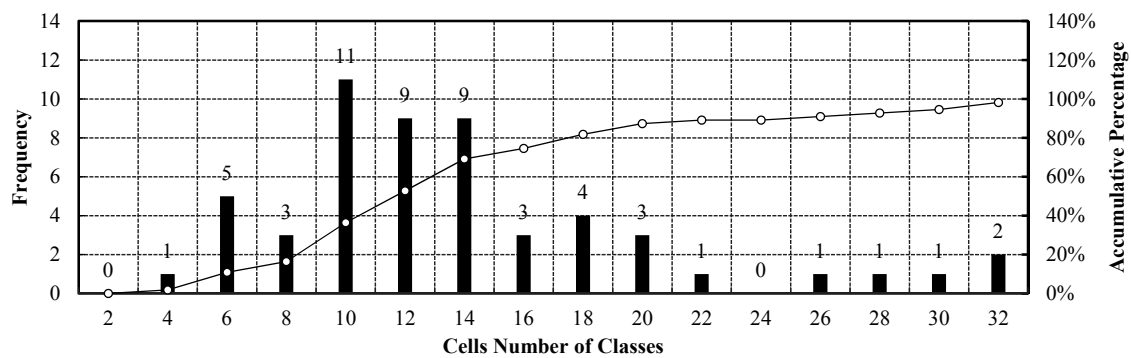


Figure 12. Number distribution of cells in each class.

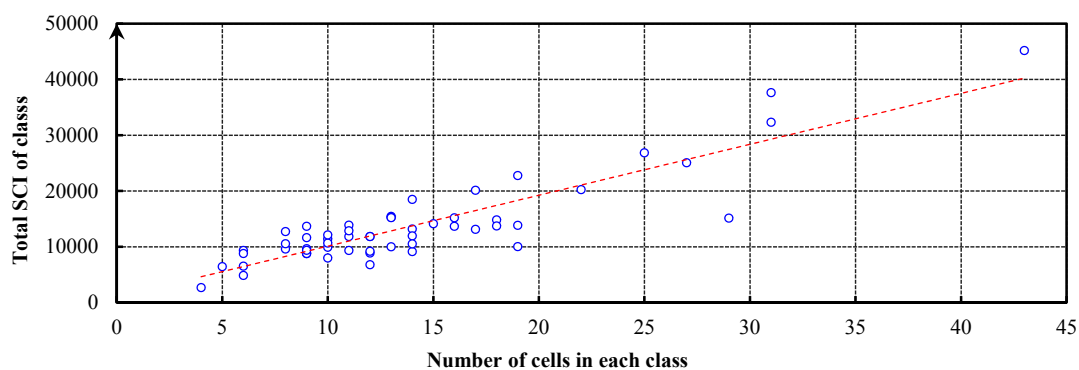


Figure 13. Cells number and total *SCI* of class.

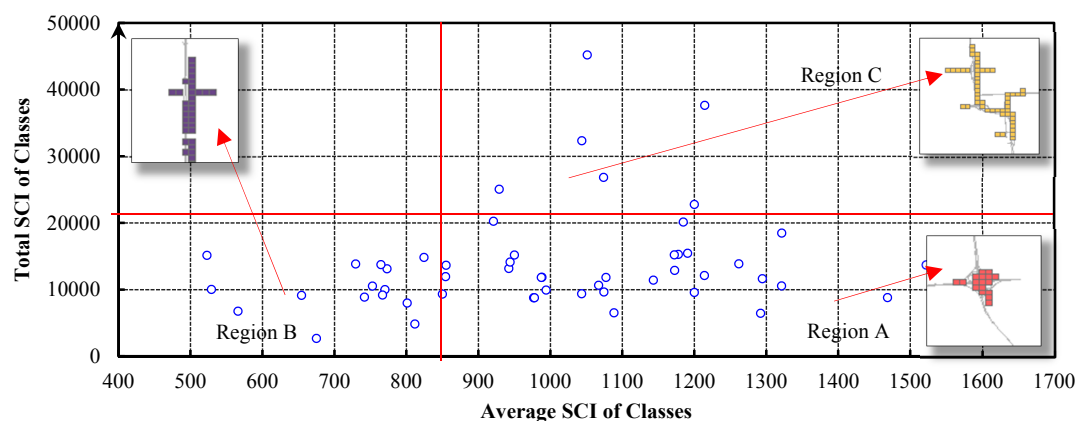


Figure 14. Average *SCI* of class and total *SCI* of class.

4. Conclusions

In order to provide a high-efficiency method for recurring traffic congestion in an urban road network, we developed a cell-based model using the taxi FCD and the DBSCAN algorithm. Taking the Beijing road network as an example, the study illustrated how to map the grid network based on identifying the proper cell size. It was found that a cell size of $100 \times 100 \text{ m}^2$ can well represent the major road network with sufficient taxi FCD samples. Then, we proposed the DBSCAN algorithm to fit the cell-based traffic congestion analyses through assigning the traffic operation performance index to the grid cells. The grid modeling results showed that the traffic congestion patterns can be categorized into three types, namely: point type, stemming from insufficient capacities at the nodes of the road network; line type, caused by high traffic demand or bottleneck issues in the road segments; and region type resulting from multiple high-demand expressways merging and connecting to each other. Through

the application of ArcGIS software, 3D maps of the road network congestion were easily generated for either dynamic operation performance visualization or location identification for different types of congestion. The proposed method would be useful for traffic management departments in order to identify the best corresponding engineering countermeasures to relieve the network traffic congestion issues. In future studies, bus FCD data and mobile sensor data can be also applied into the grid model in order to increase the sample size for each cell and the estimation accuracy of traffic operation performance. Especially, if urban POI (point of interest) data are available, the properties of land use data in cells can be integrated with the traffic analyses to find out the deep causation of congestion formation from the perspective of drivers' travel purposes and behavior features.

Acknowledgments: This work was financially supported by Chinese National 863 Project (2015AA124103), and the Fundamental Research Funds for the Central Universities (2015YJS094). The authors are grateful to anonymous reviewers for their most helpful comments.

Author Contributions: Yang Liu and Xuedong Yan conceived the study; Yang Liu and Xuedong Yan produced the theoretical material, proposed the model and solution method, and drafted the manuscript. Yang Liu, Yun Wang, and Jiawei Wu edited and coordinated the study. Jiawei Wu and Zhuo Yang helped edit the language. All authors participated in the design of the study, helped draft the manuscript, and gave final approval for publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dowling, R.; Skabardonis, A.; Carroll, M.; Wang, Z. Methodology for measuring recurrent and nonrecurrent traffic congestion. *Transport. Res. Rec.* **2004**, *1867*, 60–68. [[CrossRef](#)]
2. Wen, H.; Sun, J.; Zhang, X. Study on traffic congestion patterns of large city in China taking Beijing as an example. *Procedia* **2014**, *138*, 482–491. [[CrossRef](#)]
3. Reinthaler, M.; Nowotny, B.; Weichenmeier, F.; Hildebrandt, R. Evaluation of speed estimation by floating car data within the research project Dmotion. In Proceedings of the 14th World Congress on Intelligent Transport Systems, Beijing, China, 9–13 October 2007.
4. Leduc, G. *Road Traffic Data: Collection Methods and Applications*; JRC-IPTS Working Papers; European Commission Joint Research Centre Institute for Prospective Technological Studies: Seville, Spain, 2008.
5. Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; Ma, W.Y. Understanding mobility based on GPS data. In Proceedings of the International Conference on Ubiquitous Computing, Seoul, Korea, 21–24 September 2008; ACM: New York, NY, USA, 2008; pp. 312–321.
6. Hellinga, B.; Izadpanah, P.; Takada, H.; Fu, L. Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transport. Res. C Emerg. Technol.* **2008**, *16*, 768–782. [[CrossRef](#)]
7. Cetin, M.; Comert, G. Estimating Queues at Signalized Intersections: Value of Location and Time Data from Instrumented Vehicles. In Proceedings of the Intelligent Vehicles Symposium, Istanbul, Turkey, 13–15 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1138–1143.
8. Xu, T.; Ou, D.; Yang, Y. An ArcGIS-Based hybrid topological map matching algorithm for GPS probe data. In Proceedings of the International Conference on Transportation Engineering, Chengdu, China, 19–20 October 2013; ASCE Library: Reston, VA, USA, 2013; pp. 2421–2427.
9. Hao, J.; Zhu, J.; Zhong, R. The rise of big data on urban studies and planning practices in China: Review and open research issues. *J. Urban Manag.* **2015**, *4*, 92–124. [[CrossRef](#)]
10. Tang, L.; Huang, F.; Zhang, X.; Xu, H. Road Network Change Detection Based on Floating Car Data. *J. Netw.* **2012**, *7*, 1063–1070. [[CrossRef](#)]
11. Min, W.; Wynter, L. Real-time road traffic prediction with spatio-temporal correlations. *Transport. Res. C Emerg. Technol.* **2011**, *19*, 606–616. [[CrossRef](#)]
12. D'Andrea, E.; Marcelloni, F. Detection of traffic congestion and incidents from GPS trace analysis. *Expert Syst. Appl.* **2016**, *73*, 43–56. [[CrossRef](#)]
13. Ulm, M.; Heilmann, B.; Asamer, J.; Graser, A.; Ponweiser, W. Identifying Congestion Patterns in Urban Road Networks Using Floating Car Data. In Proceedings of the Transportation Research Board 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015.

14. Lorkowski, S.; Mieth, P.; Schäfer, R.P. New ITS applications for metropolitan areas based on Floating Car Data. In Proceedings of the ECTRI Young Researcher Seminar, Den Haag, The Netherlands, 11–13 May 2005.
15. Basnayake, C. Automated traffic incident detection with GPS equipped probe vehicles. In Proceedings of the 17th International Technical Meeting of the Satellite Division of the Institute of Navigation, Long Beach, CA, USA, 21–24 September 2004.
16. Kamran, S.; Haas, O. A Multilevel Traffic Incidents Detection Approach: Identifying Traffic Patterns and Vehicle Behaviours using real-time GPS data. In Proceedings of the Intelligent Vehicles Symposium, Istanbul, Turkey, 13–15 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 912–917.
17. Schafer, R.P.; Strauch, D.; Keplin, R. A floating-car data detection approach of traffic congestion. *Br. J. Ophthalmol.* **2001**, *64*, 211–216.
18. Yoon, J.; Noble, B.; Liu, M. Surface street traffic estimation. In Proceedings of the International Conference on Mobile Systems, Applications and Services, San Juan, Puerto Rico, 11–13 June 2007; ACM: New York, NY, USA, 2007; pp. 220–232.
19. Li, X.; Han, J.; Lee, J.G.; Gonzalez, H. Traffic Density-Based Discovery of Hot Routes in Road Networks. In Proceedings of the Advances in Spatial and Temporal Databases, International Symposium, SSTD 2007, Boston, MA, USA, 16–18 July 2007; Springer: New York, NY, USA, 2007; pp. 441–459.
20. Ghosh, B.; Smith, D.P. Customization of automatic incident detection algorithms for signalized urban arterials. *J. Intell. Transp. Syst.* **2014**, *18*, 426–441. [[CrossRef](#)]
21. Srinivasan, D.; Jin, X.; Cheu, R.L. Evaluation of adaptive neural network models for freeway incident detection. *IEEE Trans. Intell. Transport. Syst.* **2004**, *5*, 1–11. [[CrossRef](#)]
22. Gao, M.; Zhu, T.; Wan, X.; Wang, Q. Analysis of Travel Time Patterns in Urban Using Taxi GPS Data. In Proceedings of the Green Computing and Communications, Beijing, China, 20–23 August 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 512–517.
23. Bacon, J.; Beresford, A.R.; Evans, D.; Ingram, D. TIME: An Open Platform for Capturing, Processing and Delivering Transport-Related Data. In Proceedings of the IEEE Consumer Communications and NETWORKING Conference, Las Vegas, NV, USA, 10–12 January 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 687–691.
24. Janecek, A.; Hummel, K.A.; Valerio, D.; Ricciato, F.; Hlavacs, H. Cellular data meet vehicular traffic theory: Location area updates and cell transitions for travel time estimation. In Proceedings of the ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; ACM: New York, NY, USA, 2012; pp. 361–370.
25. Deng, B.; Denman, S.; Zachariadis, V.; Jin, Y. Estimating traffic delays and network speeds from low-frequency GPS taxis traces for urban transport modelling. *Eur. J. Transp. Infrastruct. Res.* **2015**, *15*, 639–661.
26. Wu, J.; Yan, X.; Radwan, E. Discrepancy analysis of driving performance of taxi drivers and non-professional drivers for red-light running violation and crash avoidance at intersections. *Accid. Anal. Prev.* **2016**, *91*, 1. [[CrossRef](#)] [[PubMed](#)]
27. Zhao, X.J.; Zhao, J.Y. Research on model of resource management for traffic grid. *Procedia Eng.* **2011**, *15*, 1476–1480. [[CrossRef](#)]
28. Bauza, R.; Gozalvez, J. Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. *J. Netw. Comput. Appl.* **2013**, *36*, 1295–1307. [[CrossRef](#)]
29. Rahmani, M.; Jenelius, E.; Koutsopoulos, H.N. Route travel time estimation using low-frequency floating car data. In Proceedings of the International IEEE Conference on Intelligent Transportation Systems, The Hague, The Netherlands, 6–9 October 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 2292–2297.
30. Talmor, I.; Mahalel, D. Signal design for an isolated intersection during congestion. *J. Oper. Res. Soc.* **2007**, *58*, 454–466. [[CrossRef](#)]
31. Scholars, S.; Szeto, S.A.; Jia, B.E.A. Research on analysis method of traffic congestion mechanism based on improved cell transmission model. *Discret. Dyn. Nat. Soc.* **2012**, *1965*, 1035–1052.

