COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING

Computer-Aided Civil and Infrastructure Engineering **32** (2017) 252–267

Mapping to Cells: A Simple Method to Extract Traffic Dynamics from Probe Vehicle Data

Zhengbing He

School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Liang Zheng*

School of Traffic and Transportation Engineering, Central South University, Changsha, China

Peng Chen

School of Transportation Science and Engineering, Beihang University, Beijing, China

&

Wei Guan

School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

Abstract: *In the era of big data, mining data instead of collecting data are a new challenge for researchers and engineers. In the field of transportation, extracting traffic dynamics from widely existing probe vehicle data is meaningful both in theory and practice. Therefore, this article proposes a simple mapping-to-cells method to construct a spatiotemporal traffic diagram for a freeway network. The method partitions a network region into small square cells and represents a real network inside the region by using the cells. After determining the traffic flow direction pertaining to each cell, the spatiotemporal traffic diagram colored according to traffic speed can be well constructed. By taking the urban freeway in Beijing, China, as a case study, the mapping-to-cells method is validated, and the advantages of the method are demonstrated. The method is simple because it is completely based on the data themselves and without the aid of any additional tool such as Geographic Information System software or a digital map. The method is efficient because it is based on discrete space-space and time-space homogeneous cells that allow us to match the probe data through basic operations of arithmetic. The method helps us understand more about traffic congestion from the probe data, and then aids in carrying out various transportation researches and applications.*

*To whom correspondence should be addressed. E-mail: *zhengliang@csu.edu.cn*.

DOI: 10.1111/mice.12251

1 INTRODUCTION

In the era of big data that have come, people no longer worry about the sufficiency of data to support researches and practical applications. Instead, how to make use of the data, which are numerous both in variety and quantity, becomes a new problem concerning researchers and engineers. It is also true in the field of transportation. Among various transportation data, probe vehicle data widely exist in cities nowadays. The probe vehicle data (or probe data) are collected by probe vehicles (or probes), which travel in the road network as regular vehicles but upload their status information to data centers every short time interval. The status information usually includes latitude, longitude, instantaneous speed, moving direction, etc. Driver's mobile phone could provide the data potentially, while taxis might be more reliable probes, because dedicated

Global Positioning System (GPS) devices installed in the taxis could communicate constantly and steadily. Although the probe data are usually treated as vehicle trajectories, the GPS points consisting of the trajectories widely scatter in the network due to the fact that the sampling intervals are usually 0.5–5 minutes, i.e., low temporal resolution.

Besides making use of the probe data to understand human mobility (e.g., González et al., 2008; Jiang et al., 2009) and to estimate travel time (e.g., Jenelius and Koutsopoulos, 2015; Li and Souleyrette, 2016), how to visualize and extract traffic conditions and dynamics from the probe data is an interesting and meaningful topic that helps people learn more about the traffic performance and further manage the traffic. To visualize the traffic conditions by using the probe data, the simplest way is to directly plot the positions of probe vehicles in spots on a map and color them based on their instantaneous speeds (Ferreira et al., 2013). However, such a highly discrete manner cannot systematically provide us with traffic dynamics that continuously evolve in time and space. Therefore, to mine more useful information, Andrienko and Andrienko (2008) introduced a mosaic diagram to show the variation of the speeds in space. In the proposed diagram, a number of mosaics were pinned on a road map, and each mosaic was employed to illustrate the temporal evolution of the traffic covered by the mosaic. Specifically in the mosaic, the x-axis was day of week, the y-axis was hour of day, and the resolution was 1 day by 1 hour in the article. Tominski et al. (2012) visualized the probe data in a hybrid two-dimensional (2D)/three-dimensional (3D) display (trajectory wall), in which 2D display presented the map in space and a time dimension (3D display) grew on the subject road to illustrate the traffic evolution. Wang et al. (2013, 2014) took a road as a whole, and illustrated the traffic of the road in 2D grid diagrams, where two axes were time of day and days, and the grids were colored according to traffic speed. Based on a polar system, Jiang et al. (2015) introduced a circular pixel graph for scalable exploration of the spatiotemporal patterns of taxi origin-destination data. The origin, i.e., the center of the circular graph, was a region selected by users from a map, and a pixel inside the graph represented the quantity of the vehicle trajectories that originated from the center and destined to the points located on the corresponding radial line. Although the visualization approaches have enriched our knowledge about the traffic, most of them fail to illustrate traffic dynamics, which propagates both in time and space, such as the propagation of traffic oscillations or so-called stop-and-go waves (Laval and Leclercq, 2010).

To identify traffic congestion and understand traffic dynamics, transportation researchers prefer a spatiotemporal traffic diagram, in which x-axis is time, y-axis is space, and the color inside (or z-axis) represents speed. The information-rich spatiotemporal diagram is a popular and powerful tool in the study and practice of transportation, such as fusing data (Treiber and Helbing, 2002; Van Lint and Hoogendoorn, 2010; Treiber et al., 2011), understanding traffic characteristics (Wilson, 2008; Kerner, 2009), and proposing and validating models (Duret et al., 2011; Ramezani and Geroliminis, 2015; He et al., 2015b). Traditionally, it is constructed by using stationary detector data, such as Chen et al. (2004), Kerner et al. (2004), Laval et al. (2009), Wieczorek et al. (2009), and Yildirimoglu and Geroliminis (2013). Few researches really focus on the construction using the probe data. Kerner et al. (2013) and Herrera et al. (2010) directly plotted the probe data with high temporal resolution into a time-space plane, and the traffic dynamic was roughly visualized. A premise of doing this was that they mastered the map-matching and Geographic Information System (GIS) techniques and had a digital map for the road network, so that the probe data can be matched on the corresponding road.

However, there are two obstacles hindering us from making use of the map-matching and GIS techniques to construct the spatiotemporal diagram. The first one is the map-matching and GIS techniques. Map-matching is a well-known challenging task (Quddus et al., 2007; Quddus and Washington, 2015), in particular for the probe data with low temporal resolution (Hunter et al., 2014). Also, it may be difficult to acquire them in a short time, although related techniques and software widely exist. In particular, these techniques are rather like complicated tools than real interests for the transportation engineers and researchers who are less relevant to GIS research. Moreover, it should not be ignored that the commercial GIS software is usually expensive. The second one may be worse, i.e., no digital map that allows researchers and engineers to match the probe data. Our case is like this. A company provided us a large number of probe data in months, but was not willing to share the map to protect the intellectual property. Other open-source digital maps may not well match these data. Thus, it seems impossible to analyze the data without the aid of a digital map. We believe that the lack of a proper map happens somewhere else, in particular for the data that are not dedicated to academic research, when we cannot easily obtain a map, or in the countries where geographic information is restricted.

Therefore, to visualize traffic dynamics completely from the probe data and accomplish it in a simple manner, this article proposes a simple but efficient mapping-to-cells method to construct the spatiotemporal traffic diagram for a freeway network. The method is simple,

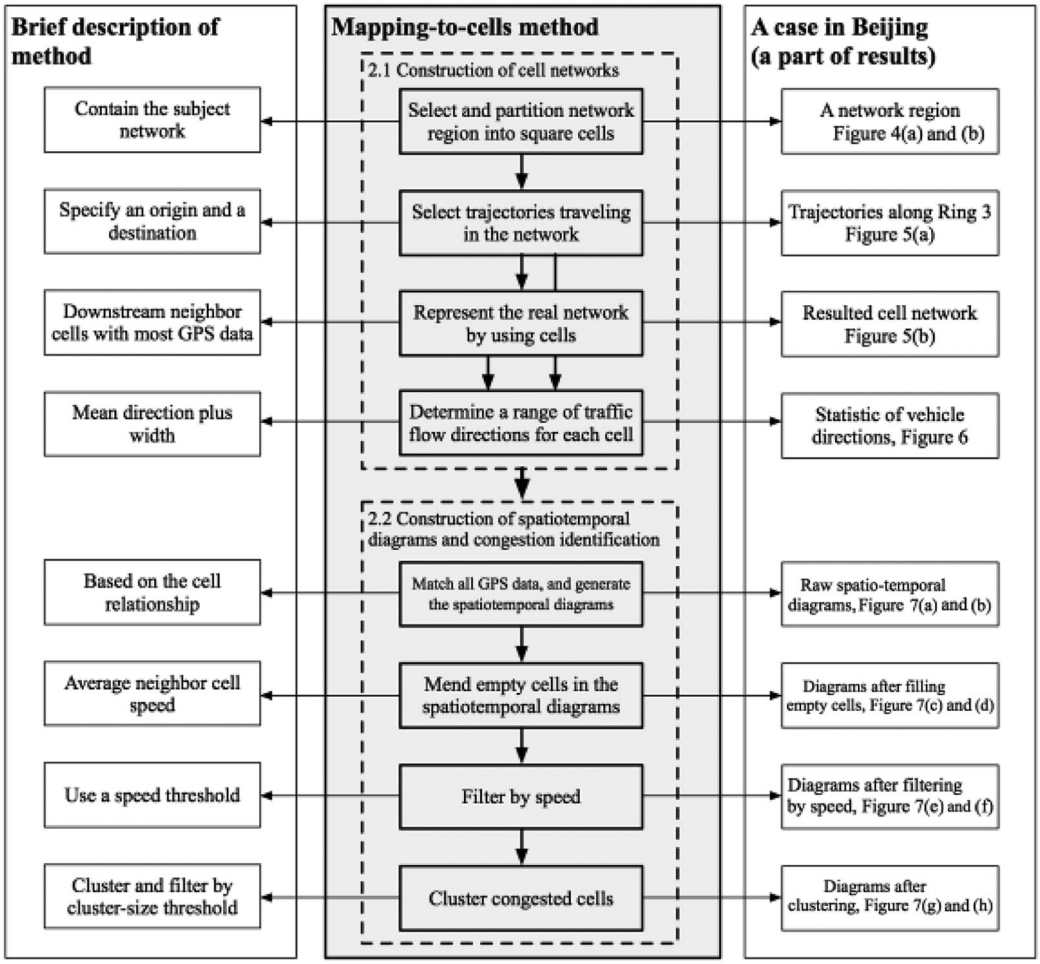

Fig. 1. A flowchart of the mapping-to-cells method.

because it is purely based on the probe data and without making use of any additional tool, such as GIS software or a digital map. The researchers and engineers who are not familiar with the map-matching and GIS techniques could master it in a short time. The method is efficient and potentially suitable for online applications, because it transforms the continuous network space into discrete and homogeneous cells, and we can match GPS points with the cells through simple calculation. The efficiency rather than accuracy may be more meaningful to process big data. By taking the urban freeway in Beijing, China, as a case study, the mapping-to-cells method is validated and its usefulness is demonstrated. To the best knowledge of the authors, it is the first time that the spatiotemporal traffic diagrams of Beijing urban freeways are presented in such a resolution, which is thanks to the mapping-to-cells method.

The rest of the article is organized as follows: Section 2 addresses the mapping-to-cells method, including construction of a cell network, determination of traffic flow direction in cells, and construction of the spatiotemporal diagram and congestion identification. Section 3 demonstrates the method in a case of an urban freeway in Beijing, China, where the method is validated and stochastic congestion maps are constructed besides the illustration of constructing the spatiotemporal traffic diagrams. A conclusion is made at last. Moreover, we summarize the structure of this article in Figure 1, which includes the procedures of the mapping-to-cells method, the brief description of each procedure, and the corresponding results in the case of Beijing urban freeway.

2 MAPPING TO CELLS: THE METHODOLOGY

2.1 Construction of cell networks

This subsection introduces the conversion from a real freeway network to a network composed of square cells. Note that sometimes, one refers to the freeway with two

opposite directions as a single freeway; here, we treat each direction of the freeway as a separate freeway.

First, we select a network region containing the subject freeway and partition the region into homogeneous square cells. Generally speaking about the selection of the cell size, the denser the GPS data are, the smaller the cell size that we can choose is. If we choose a large size, like over 500 m, we should make sure that a cell does not cover two parallel freeways with the same directions, i.e., the road density should be considered. If a cell covers two or more parallel freeways with the same directions, the method will fail because the proposed method cannot distinguish the traffic on such freeways. If we choose a small size, like tens of meters, we may not have such dense GPS data capable of falling into most of the cells during a short time interval like 1 or 2 minutes, i.e., the density of GPS data should be considered. In addition, the cell should not be smaller than the width of a freeway. For a method of analyzing empirical data, it is difficult to give an optimal value or a method to determine the optimal value under the condition of lacking complete data. However, considering various factors, we recommend a cell between 50×50 m^2 and 200×200 m^2. In the Beijing case, we will show that the proposed method is not very sensitive to the selections.

Second, to represent a freeway by using the cells, we need to select sample trajectories of vehicles that travel on the freeway. The trajectories will clearly make the freeway emerge under the condition without a digital map. To the end, we first plot all GPS points within a time interval on a latitude–longitude plane, where a basic image of the entire network emerges. With the visual aid of the image, we could specify an origin and a destination for the freeway. Then, any trajectory that travels between them and meanwhile moves along the freeway is qualified to select. Note that the trajectories do not have to start and end within the origin and destination; passing them is accepted. Moreover, the trajectories do not have to appear at the same time interval, because we only need them to specify the freeway in space. If the temporal resolution is low, we may select more trajectories, and the GPS points coming from different trajectories but belonging to the same freeway will make the network emerge. It is usually enough if the number of the selected trajectories can make sure that there is at least a GPS point within every distance equal to the cell length. If the freeway is so long that few trajectories can cover the entire freeway, we can first split the freeway into subfreeway, then select trajectories separately, and combine all the qualified trajectories at the end.

At last, we represent the real freeway or network using the square cells. Since all cells are homogeneous squares, it is obvious that there are eight neighbors for

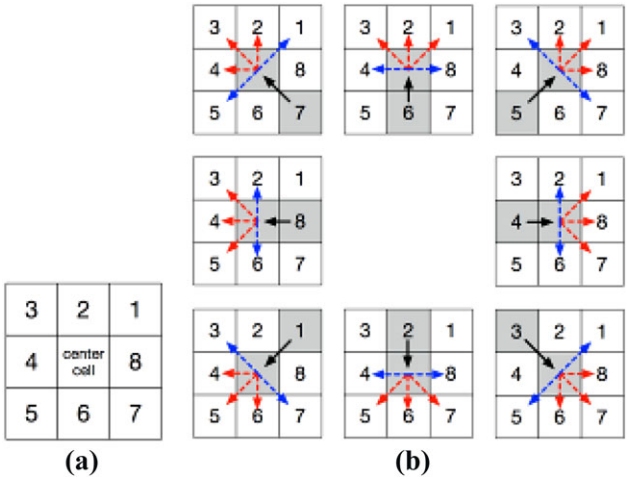

Fig. 2. Basic of a cell network: (a) a cell and its neighbor cells and (b) searching direction in mapping a freeway network to a cell network.

a cell in center, which can be numbered such as shown in Figure 2a. To depict the real freeway, we further define the following downstream and upstream relationship for the center cell: Given an upstream cell, three downstream cells and two lateral cells of the center cell are taken as the potential downstream cells (PDCs) of the center cell. The three downstream cells are defined as level-1 PDC, and the two lateral cells are defined as level-2 PDC. As shown in Figure 2b and taking the case at the upper left corner of the figure as an example, neighbor 7 is the upstream cell of the center cell, and neighbors 1–5 are the PDCs of the center cell, where neighbors 2–4 are level-1 PDC and neighbors 1 and 5 are level-2 PDC. It is obvious that there are total eight possible cases.

Based on the fact that vehicles only travel on roads, it is reasonable to believe that the real downstream cell will contain more GPS points than the other PDC. Thus, we select the cell in level-1 PDC that contains most GPS points coming from the sample trajectories. If there is no GPS point in any level-1 PDC, we then check level-2 PDC and select the cell in level-2 PDC that contains most GPS points. Selecting in two levels is to give a priority to the forward growth of the cell network, i.e., first select in level-1 PDC, because most urban freeways are usually straight, or in the curved section of a freeway, the direction changes are smooth and gradual (unlike a sharp turn at an intersection). Retaining level-2 PDC as the second option is to ensure that we can find cells to depict sharp curves of a freeway.

Now, after initially specifying a cell overlaying the freeway (denoted by n_0) and its downstream cell (denoted by $n_0 + 1$), we can determine the downstream

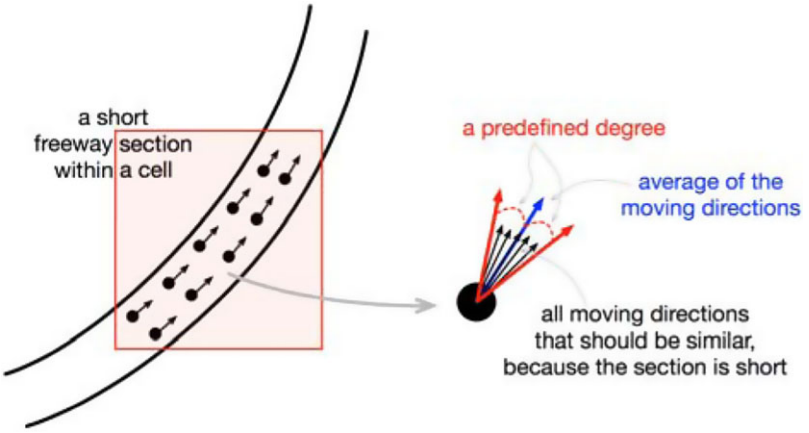

Fig. 3. An illustration of specifying the traffic flow direction for a cell.

cell $n_0 + 2$ from the PDC of the downstream cell $n_0 + 1$ following the above rule of selecting from the PDC. Repeat the process until the entire freeway is represented by the cells. It is noted that selecting the sample trajectories in advance is indispensable to guarantee the effectiveness of the process. Moreover, this is a one-shot automatic process, i.e., we only need to do the process once, and then obtain the cell network to incorporate the large amount of probe data and other applications.

Although we have already had the cell network, it is common that a cell contains GPS data in two opposite directions when it geographically covers two freeways with opposite directions, and it may even contain several different flows when being close to ramps or interchanges. To purify the GPS data and only retain those belonging to the subject freeway, we specify a range of traffic flow directions for every cell as follows (also see Figure 3 for an illustration). First, select all GPS points falling into the cell from the sample trajectories traveling on the freeway. Note that these data completely belong to the freeway because they come from the sample trajectories. Then, simply take an average of the moving directions of these data points as the center of the direction range of the cell, and extend it to a range by adding a predefined degree as the width of the range. The fact that allows us to do this is that in a short freeway section, such as a section of 50–200 m long, all passing vehicles usually maintain similar moving directions, which we will further validate in the following case study.

2.2 Construction of the spatiotemporal traffic diagram and congestion identification

After having the cell network and the traffic flow directions pertaining to each cell, we are now ready to construct the spatiotemporal traffic diagram for a freeway. Given a time interval, we match all GPS points through their longitude and latitude with the regions of the cells in the cell network, and compute the average speed of all GPS data falling into the range of the traffic flow directions, which is the aggregated traffic information of the cell during the time interval. According to the upstream and downstream relationship of the cells in the cell network, the spatiotemporal diagram can be constructed, where we set x-axis to be the time of day, y-axis to be cells, and color the diagrams according to the speed in cells or set z-axis to illustrate the speed.

It is noted that since all cells are homogeneous squares, which cell a GPS point belongs to can be directly and rapidly determined by simple calculation, i.e., first converting the longitude and latitude to the relative position in the network region, and then dividing the position by the size of the cell. Otherwise, if the cells are in different shapes or in inhomogeneous sizes, we have to compare the GPS point with all cells in a loop, which is obviously time-consuming. Thus, the computational efficiency should be high determined by the discrete nature, which is meaningful when we process a large amount of data in the era of big data.

The spatiotemporal traffic diagram is composed of time-space cells, whereas the cells in the cell network correspond to 2D space. To distinguish the two types of cells, the rest of the article denotes the cells in the spatiotemporal diagrams by using "ST cells" and denotes the cells in the cell network by using "CN cells."

Although the spatiotemporal diagram has been constructed, not all ST cells, in practice, can be filled with at least one GPS point, in particular for the uncongested ST cells where GPS points are few due to low traffic density. To mend these empty ST cells, we simply take the average of the traffic conditions (if available) of all immediate neighbors (eight in total) as the traffic condition of the empty ST cells. The way is feasible

because of high relevance among neighbor ST cells. Another prerequisite of repairing empty ST cells in this way is that there are not too many empty ST cells. Thus, a certain amount of GPS data is needed by the method.

Moreover, to identify traffic congestions, we filter the spatiotemporal diagram using a speed threshold, i.e., remove all ST cells with the speed greater than the threshold, and only retain congested ST cells. According to the fundamental diagram in the traffic flow theory (see chapter 4 in Daganzo, 1997, for example), the uncongested and congested traffic states are usually separated by critical speed. Thus, filtering by a speed threshold is like transforming the traffic conditions into binary uncongested (removed from the diagram) and congested (reserved in the diagram) states.

To further purify the congestion and expose major traffic congestion, we cluster these congested ST cells as follows: combine all neighboring congested ST cells as congested clusters, and remove the clusters with the size smaller than a cluster-size threshold. It is obvious that more severe congestions will be retained with the increase of the threshold. After the step, only major congestions are left in the spatiotemporal cluster diagram, and the relevant bottlenecks will be clearly exposed.

3 A CASE OF BEIJING URBAN FREEWAY

3.1 Beijing urban freeway and the probe vehicle data

Beijing is the capital of China and one of the largest cities in the world. At present, four two-way urban ring roads (urban freeways or expressways), i.e., rings 2–5, enclose the urban area of Beijing, and are linked by eight connecting urban freeways. The lengths of rings 2–5 are 33, 48, 65, and 98 km, respectively. All these urban freeways are surrounded by auxiliary roads, and the auxiliary roads and urban freeways are connected by exits and entrances. Although traffic signal control that is quite common at intersections does not exist, there are a large number of various types of interchanges for merging or diverging traffic. The numbers of the interchanges are 38, 48, 53, and 47 for rings 2–5, respectively. Each interchange usually corresponds to a pair of exit and entrance. On average, there is a pair of exit and entrance every 1 km, which are all potential bottlenecks. For more detail about the ring roads in Beijing and the traffic flow characteristics, one can refer to Zhao et al. (2009), He et al. (2015a), and Yu et al. (2015), for example.

The total population in Beijing had been over 21 million by 2013. To serve the giant number of people, the city holds about 10,000 taxis. About every 1 minute, each of these taxis uploads a set of instantaneous information, such as its position, moving direction, speed, being occupied, or not occupied by passengers. These data are well-qualified probe data providing on-road traffic information. In this article, we demonstrate the mapping-to-cells method by using these GPS data collected on 30 workdays (6 weeks) in the first 2 months in 2015. Only the taxis that were occupied by passengers are considered here, because the behavior of these taxis is closer to that of regular vehicles. It is noted that those GPS data are based on GCJ-02 coordinates, and those online open-source maps are usually based on WGS-84 coordinates (Restrictions on Geographic Data in China, 2016). If we put GCJ-02 based data on a WGS-84 map, the offsets between the two coordinates could result in 100–700 meter errors, which prevents us from directly utilizing those online maps. To save the space, we only select ring 3, one of the most important roadways in the city, as a study case. The results of other rings and connecting urban freeways are similar in terms of demonstrating the method. Ring 3 contains three lanes, respectively, in the opposite directions, i.e., clockwise (CW) and counterclockwise (CCW) directions, and the width of each lane is 3.75 m. The speed limit is 80 km/h.

3.2 Construction of spatiotemporal traffic diagrams for Beijing urban freeway

Following the procedures introduced in Section 2, we first select a 16×16 km^2 central region containing ring 3; see Figure 4 for the basic information about the region and the corresponding probe data. Figure 4a displays the region in an online map. Figure 4b presents the instantaneous positions of all taxis that are occupied by passengers, where the major network emerges basically. Based on the visualization, we can determine the network region and even distinguish major roads. Figure 4c presents the variations of the numbers of the GPS data within a day, which sustains around 6,000 data points during the daytime. Figure 4d shows daily changes of all the probe data during the 6 weeks.

We partition the 16×16 km^2 region into 160×160 CN cell[2], i.e., each CN cell is a 100×100 m^2 square. Then, select the sample trajectories traveling along the CW and CCW directions of ring 3, respectively. We have roughly seen ring 3 in Figure 4b, and with its visual aid, it is easy to select an origin and a destination on ring 3 for specifying the sample trajectories. Since the ring is quite long and few taxis completely circuit for an entire round, we arbitrarily set eight pairs of origins and destinations for two directions of the ring. See Figure 5a, where the blue rectangles are the selected pairs of origins and destinations. Based on the pairs, we can obtain and plot all trajectories traveling

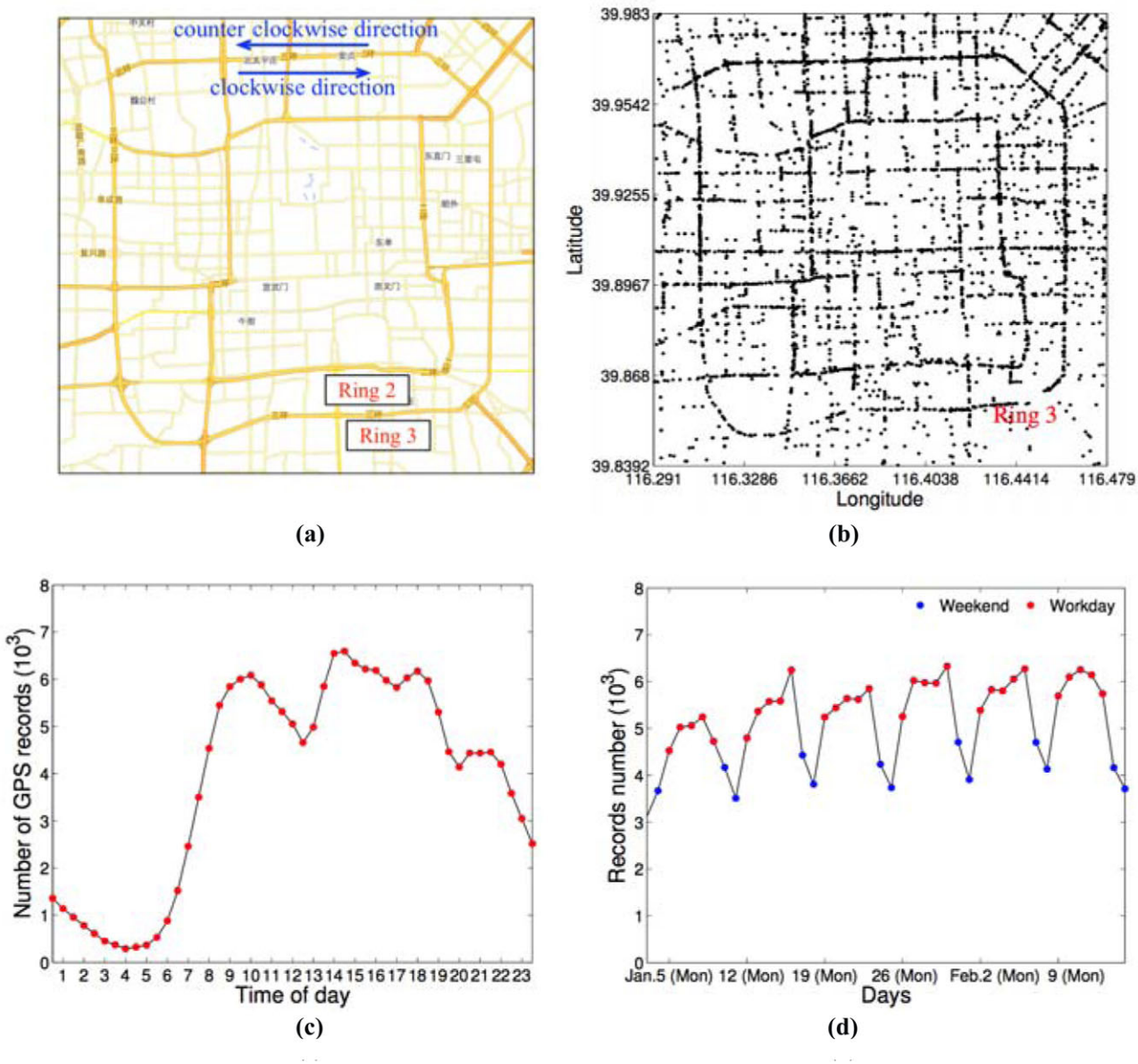

Fig. 4. A glance of the taxis occupied by passengers in Beijing urban area: (a) a network region containing ring 3 in Beijing, (b) positions of all GPS points at 8:00 on January 15, 2015, (c) number of the GPS points on January 15, 2015, and (d) daily average number of the GPS points during the period from 8:00 to 20:00.

between each of them. After manually removing all trajectories whose middle parts clearly deviate from ring 3, the remaining trajectories are all qualified. Although there are usually a lot of qualified trajectories, it is not necessary to involve all of them. In the case we select 30 trajectories, which make sure that there is at least a GPS point in every 100 m.

Based on the sample trajectories, the cell network of ring 3 can be constructed following the previously proposed two-level rule; see Figure 5b for the results. In Figure 5b, cells B and C are two of the level-1 PDC of cell A, the proposed method selects cell C as the real downstream cell of cell A, because there are more GPS points inside cell C than those inside cell B. When we construct the spatiotemporal diagram in subsequence, only the data falling into cell C will be used, and the data into cell B are discarded. Most of the data are kept by the method (i.e., those in cell C in the example), although a part of data (i.e., those in cell B) is wasted as a trade-off of the simplicity.

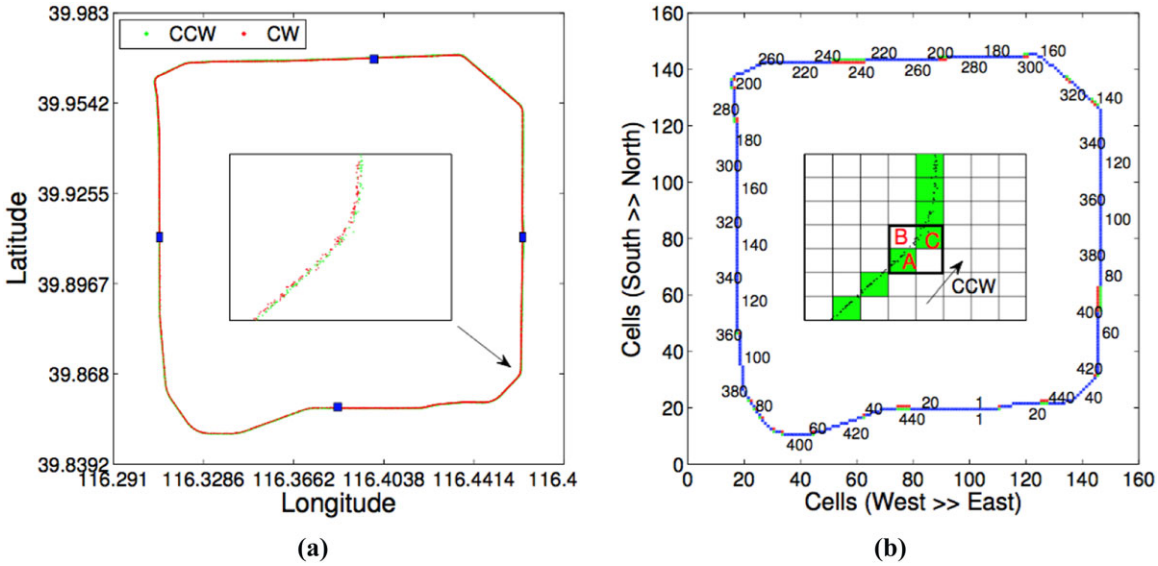

Fig. 5. The demonstration of constructing a cell network for ring 3 in Beijing: (a) selected sample trajectories along ring 3 and (b) constructed cell network. The number inside (outside) surrounding a ring indicates a relative location along the CW (CCW) direction of the ring. The green color indicates CCW cells, the red indicates CW cells, and the blue indicates the common cells. The subfigure inside illustrates a relationship between the cell network and the sample trajectories.

To validate the method of determining traffic flow directions in cells, Figure 6 presents the statistics of all the directions of the GPS points coming from the sample trajectories and falling into the cells. It can be seen that the difference between the maximum value and the mean value of all flow directions in a cell, the difference between the mean value and the minimum value, and the standard deviation of all flow directions are usually quite small for straight road sections, and are less than 20° or 30° for curve sections. It means that the traffic flow directions of the GPS points contained by a cell usually centralize in the mean value of all directions, and the deviations are small. Thus, the proposed method is feasible, i.e., taking the average as a center and extending it to a range by adding a width. According to the results shown in Figure 6, we set the width of the traffic flow direction to be 30° in this case.

After constructing the cell network and determining the traffic directions in cells, we present the spatiotemporal traffic diagrams for the two directions of ring 3 on a randomly selected workday. In constructing the diagrams, we select 2 minutes as the time interval to aggregate all the GPS points in CN cells, and all available GPS data are employed. See Figure 7, where the first-row subfigures are directly constructed from the GPS points in CN cells, the second-row subfigures are the plots where empty ST cells are mended, the third-row subfigures are filtered by a given critical speed of 20 km/h, and the forth-row subfigures present the clusters containing over 50 congested ST cells. To the best knowledge of the authors, it is the first time that the spatiotemporal traffic diagrams of Beijing urban freeways are presented in such resolution, which is thanks to the mapping-to-cells method.

From these diagrams, we can observe abundant information about the traffic on the urban freeway in Beijing, such as pinned localized cluster and oscillating congested traffic. The pinned localized cluster is defined as a single localized standing traffic wave "pinned" at a bottleneck, and the oscillating congested traffic is defined as an extended congested zone with moving traffic waves (chapter 17 in Treiber and Kesting, 2013). Zoom in the regions marked in Figures 7b and d are presented in Figure 8. It is clear that the congestion patterns are consistent with the definitions, and the speed of the moving traffic waves is about −17 km/h constantly. Alternatively, the observed traffic structure can be described as a congested pattern consisting of synchronized flow and wide moving jams (Kerner, 2009).

To better understand the case, we measure the mean, the standard deviation, and the total number of all speeds in every cell that contains at least one GPS point, and calculate the correlation coefficients between each two of them. The correlation coefficients for the CW (CCW) direction are −0.1621 (−0.1555) between the mean and the standard deviation, −0.3799 (−0.3429) between the mean and the total number, and 0.2873 (0.2702) between the standard deviation and the total number, respectively. They all indicate weak relationship among these indicators. It is noted that the negative

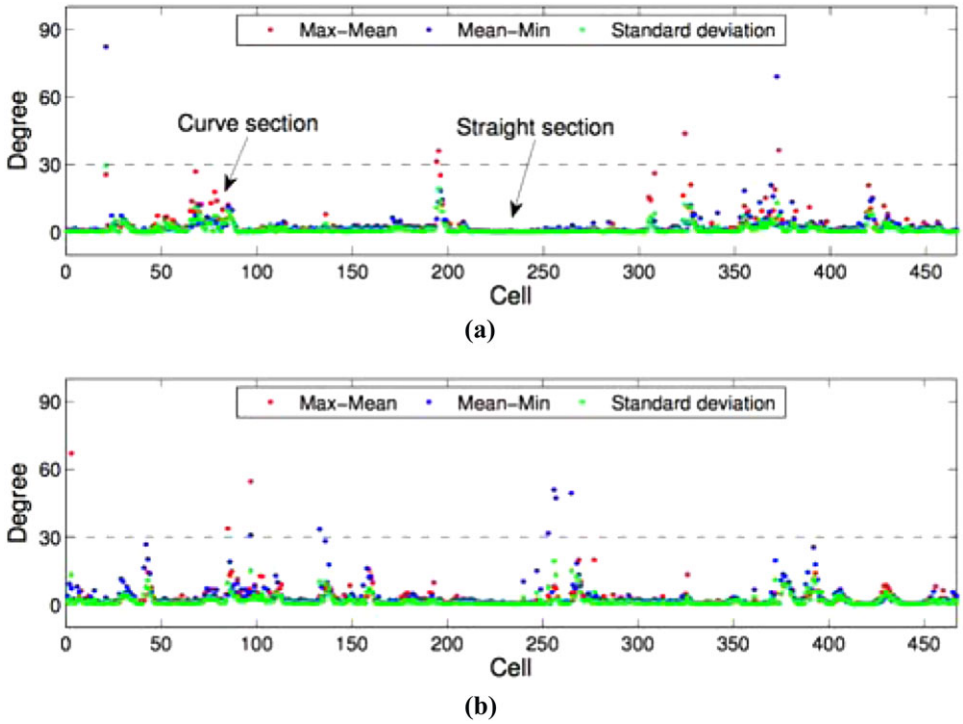

Fig. 6. Flow directions in all cells: (a) CW direction and (b) CCW direction. Note that "Max," "Min," and "Mean" indicate the maximum value, the minimum value, and the mean value of the flow directions in a cell, respectively.

correlation between the mean and the total number is due to the fact that more probe vehicles usually appear in traffic congestions. The positive correlation between the standard deviation and the total number may be resulted by the ST cells containing few GPS points. For example, the standard deviation will be zero if there is only one vehicle in the ST cell, and such cells will enforce positive correlation.

To see more detail, we plot the empirical cumulative distribution functions (ECDFs) of these standard deviations and the total numbers in Figure 9. The ECDF generally reflects the distribution or frequency of different values of a variable. Since the correlations are weak, we plot them separately. It can be seen from Figure 9a that the standard deviation larger than 20 km/h is only about 10%. A large percentage of the standard deviations are equal to zero, which is caused by the ST cell containing few GPS points. Figure 9b shows that there are less than 20% empty ST cells, and a number of the cells filled by two to six GPS points, i.e., there are usually two to six GPS points every 100×100 m^2 during 2 minutes.

3.3 Sensitivity analysis

This subsection demonstrates the influence of selecting CN and ST cells with different sizes. In general, the size of the CN cell reflects the magnitude of aggregating data in space, and the time interval of the ST cells reflects the magnitude of aggregating data in time. Enlarging the CN cell or the time interval of the ST cells is to manually involve and average more data in space or in time. Thus, employing smaller CN cells or shorter interval ST cells can make the diagrams closer to real data due to less data being averaged. However, one drawback of the reduction is resulting in more empty cells. Then, effectively mending empty ST cells is more crucial. Basically, the proposed method, which is to average data in space and time in essence, shows satisfactory performance.

To intuitively show the influence, we construct the spatiotemporal diagram with different settings, and focus on the same region shown in Figure 8 in order to make comparisons; see Figure 10. Figures 10a and b are constructed with the time interval of 0.5 minutes and the CN cell size of 100×100 m^2. Compared with Figure 8, it can be seen that there are more empty ST cells in Figure 10a, whereas the mending method successfully fills the empty ST cells (see Figure 10b). Figures 10c and d are constructed with the time interval of 2 minutes and the CN cell size of 200×200 m^2. Their differences from Figure 8 are not obvious, although we can still see that they are blurrier than Figure 8 by carefully comparing the oscillating traffic. Meanwhile, there

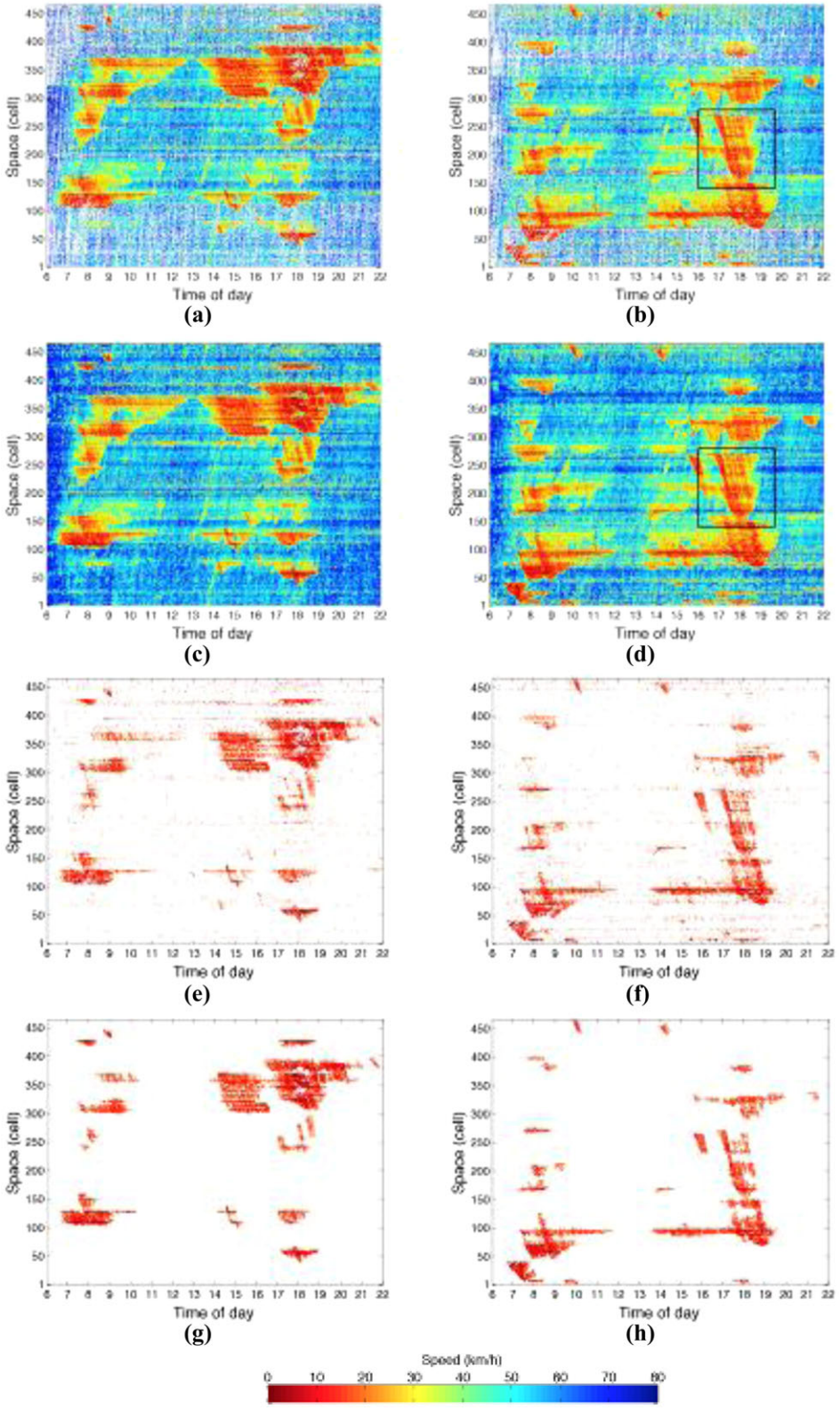

Fig. 7. Spatiotemporal traffic diagrams on January 15, 2015: (a), (c), (e), and (g) CW direction; (b), (d), (f), and (h) CCW direction; (a) and (b) original spatiotemporal diagrams; (c) and (d) after filling empty cells; (e) and (f) after filtering by speed; and (g) and (h) after clustering congested cells.

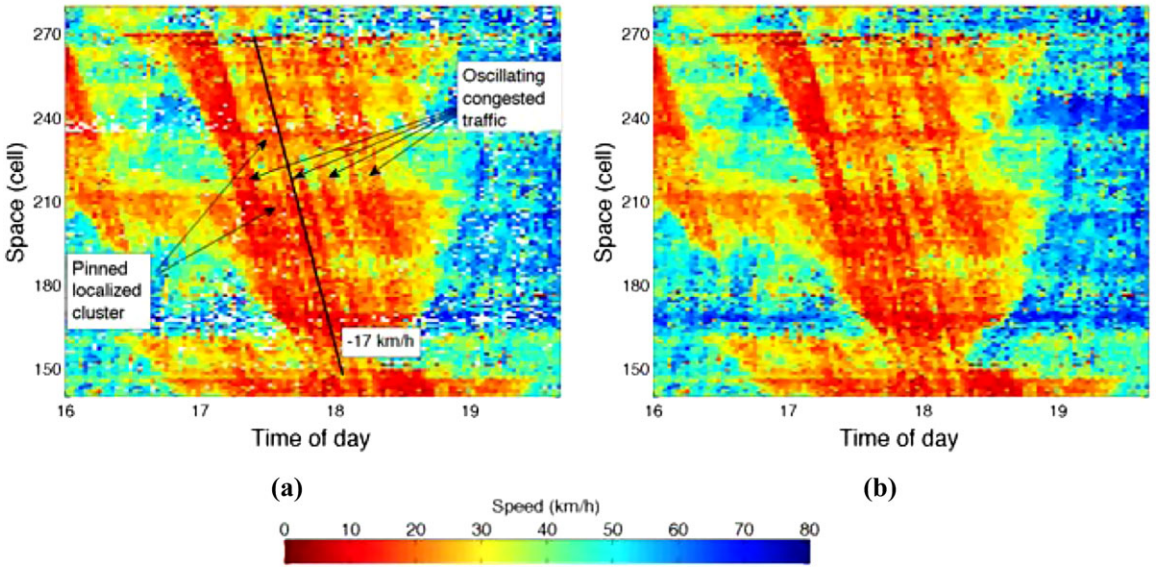

Fig. 8. An observation of the typical traffic congested patterns on Beijing urban freeway. The time interval of ST cells is 2 minutes: (a) original diagram and (b) diagram after empty ST cells are filled.

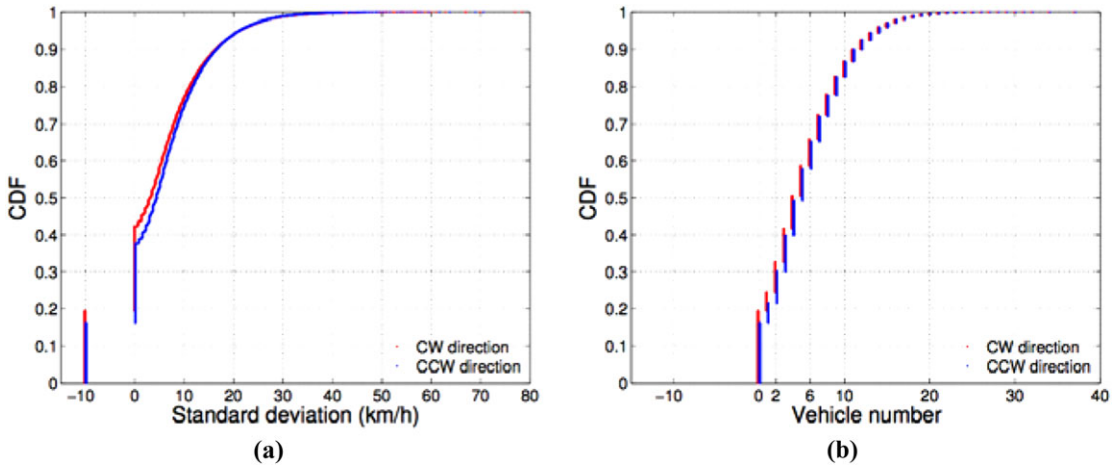

Fig. 9. Comparisons of GPS points in cells based on empirical cumulative distribution functions: (a) standard deviation of all speeds in all cells, where "−10" denotes empty cells and (b) GPS point number in all cells.

is almost no empty cell in Figure 10c because the cell with the larger size involves more GPS data. Even when the CN cell is enlarged, the congestion and bottlenecks are also well visualized. However, it is noted that although traffic oscillations can be observed roughly, the resolution may not be high enough to accurately analyze these waves. This magnitude is determined by the density of the probe vehicles and sampling frequency of the GPS device.

To better compare these diagrams, Figure 11a presents the ECDF for the speeds in Figures 8b, 10b, and 10d, respectively. It can be seen that the shapes of the ECDF are similar, in particular when the cell sizes are the same (i.e., 100×100 m^2); see the red and green curves. In contrast, the blue curve (i.e., 200×200 m^2, 2 minutes) deviates from the red and green ones. To see the difference more clearly, we first fit the ECDF curves by using Kernel density estimation, which is a popular nonparametric method to estimate the probability density function of a random variable. Then, calculate the difference between each two of the three fitting curves, and present the results in Figure 11b. It can be seen that the largest deviation of 0.04 between the fitted blue curve (i.e., 200×200 m^2, 2 minutes) and other fitted curves occurs at the speed greater than 50 km/h. Another relatively large deviation of −0.03 occurs at the

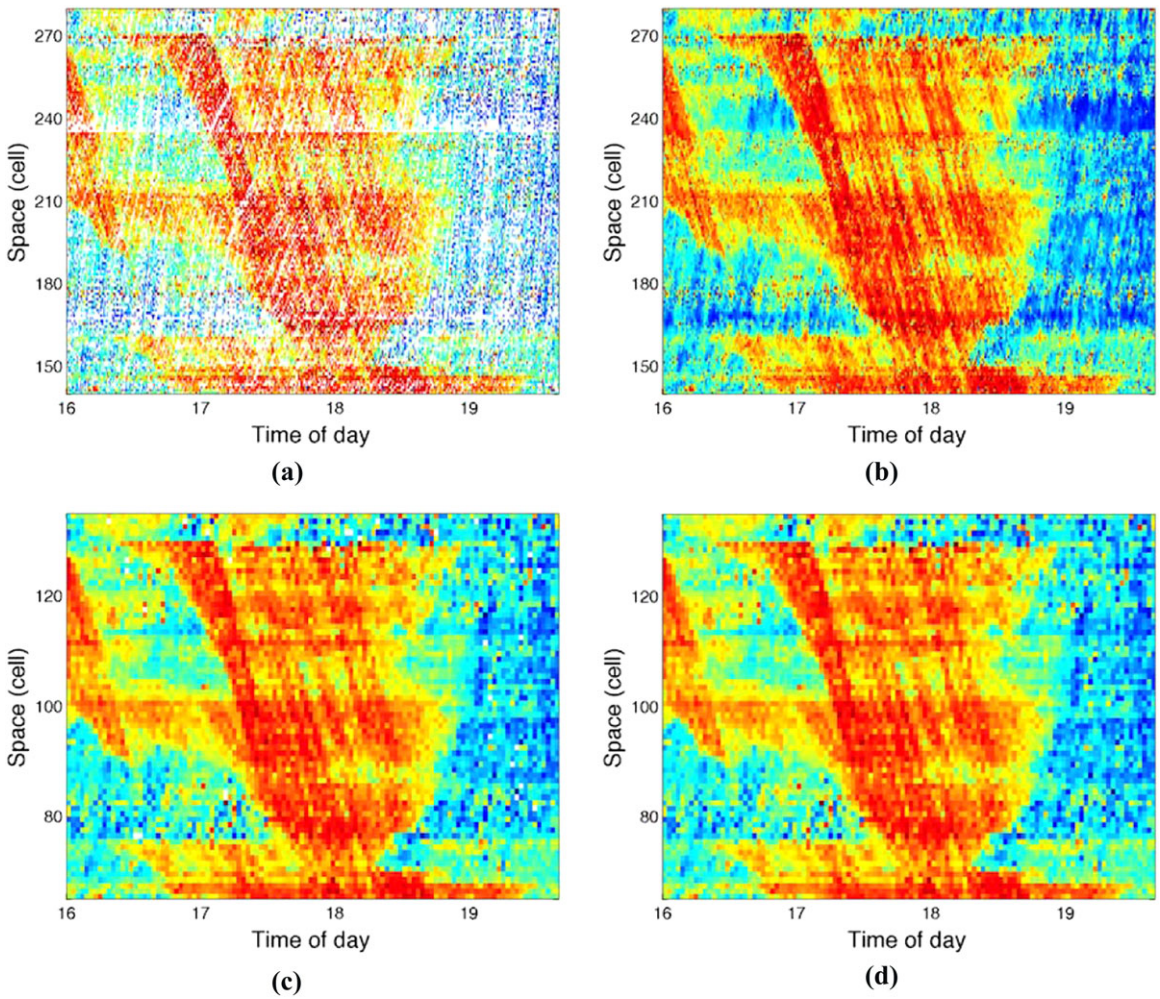

Fig. 10. Spatiotemporal traffic diagrams constructed with different settings: (a) and (b) diagrams with time interval of 0.5 minutes and cell size of 100 × 100 m^2; (c) and (d) diagrams with time interval of 2 minutes and cell size of 200 × 200 m^2; (a) and (c) original diagrams; and (b) and (d) diagrams after empty ST cells are filled.

speed of 15 km/h. The deviations occurring at low and high speeds is due to the fact that larger CN and ST cells usually involve more GPS data and extreme values (i.e., low and high speed) may be smoothed.

In general, Figures 8, 10, and 11 demonstrate good similarity in the spatiotemporal diagrams constructed with different settings. It implies that the mapping-to-cells method is not very sensitive to time intervals and cell sizes in this case.

3.4 Comparisons with instantaneous GPS speed

Except the above analysis, we randomly select 10 trajectories out of the 30 sample trajectories traveling along the ring, and compare all the instantaneous speed with the aggregated speed in their corresponding ST cells. Figure 12 presents all of the comparisons, where we sort them by the instantaneous GPS speed to make a clear demonstration. It can be seen that although errors occur, the speeds in ST cells (blue and red spots) are able to approximate to the instantaneous speeds (black spots), and the low speed can also be captured by the clusters (red spots). The comparisons are between average speed (speed in cells) and single speed (instantaneous GPS speed from the trajectories), and the average is taken within 100 m × 2 minutes, during which the traffic may change. Thus, the speeds in cells (blue and red spots) scatter around the GPS speed (black spots). However, capturing the trend is more important for a taking-average method. It is also noticed that when the instantaneous speeds are low (or high), such as lower than 20 km/h (or higher than 60 km/h), the speeds in ST cells are usually higher (or lower). This can be attributed to the fact that the speeds in ST cells

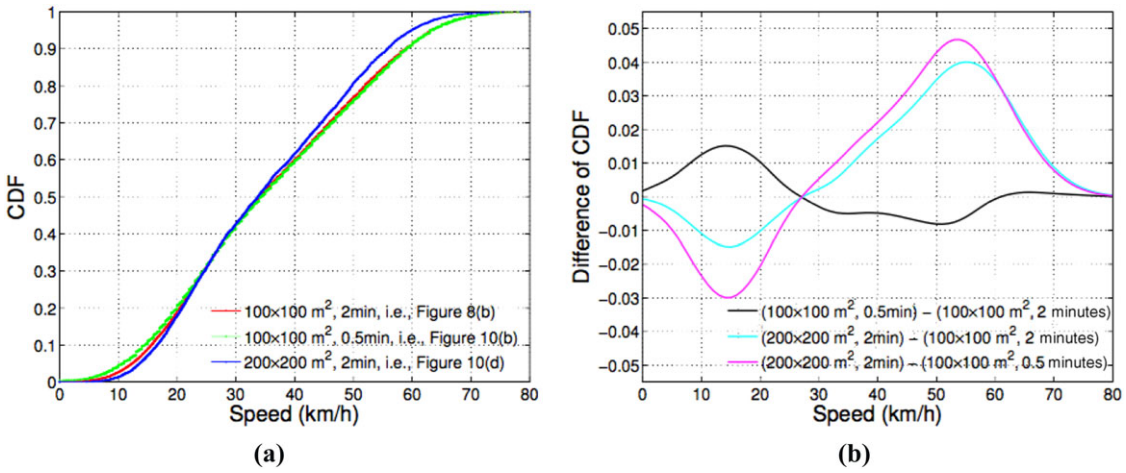

Fig. 11. Comparisons of the empirical cumulative distributions built by using the speeds in all cells of the zoom-in spatiotemporal diagrams after the empty cells are filled: (a) the empirical cumulative distributions for the speeds in Figures 8b, 10b, and 10d; and b the difference between each two of the three distributions.

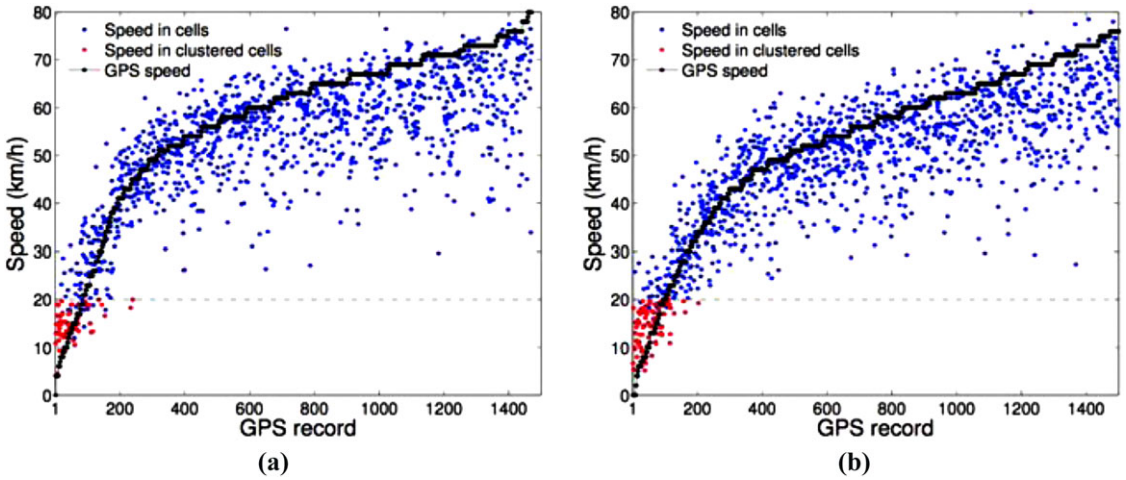

Fig. 12. Comparisons between the instantaneous GPS speed and the aggregated speed in ST cells. To make the presentation clear, we sort the comparison by the GPS speed: (a) CW direction and (b) CCW direction.

Table 1
Statistical results of the comparisons between the instantaneous GPS speed and the aggregated speed in ST cells

	Total number	*MAE*	*RMSE*	*MAPE*	*PCCC*
CW direction	1,470	5.68	8.78	13.1%	97.0%
CCW direction	1,515	6.10	8.77	15.0%	97.2%

are aggregated values, which smoothens extreme single values.

Table 1 presents the corresponding statistical results. It can be seen that both mean absolute error (MAE) and root mean square error (RMSE) are lower than 10, and mean absolute percentage error (MAPE) is about 15%. It is acceptable for a comparison between the instantaneous speed and the aggregated speed in 100 m $\times$ 2 minutes. Moreover, we propose an indicator called percentage correctly captured by clusters (PCCCs), i.e., the percentage that the instantaneous speed is smaller than the speed threshold when the GPS point falls in a cluster ST cell. It can be seen that correctly captured percentages are as high as about 97%, which means the cluster is able to well reflect the congestion.

3.5 Stochastic congestion maps

To further validate and utilize the method, this subsection constructs stochastic congestion maps for the two directions of ring 3, and analyzes the daily traffic and

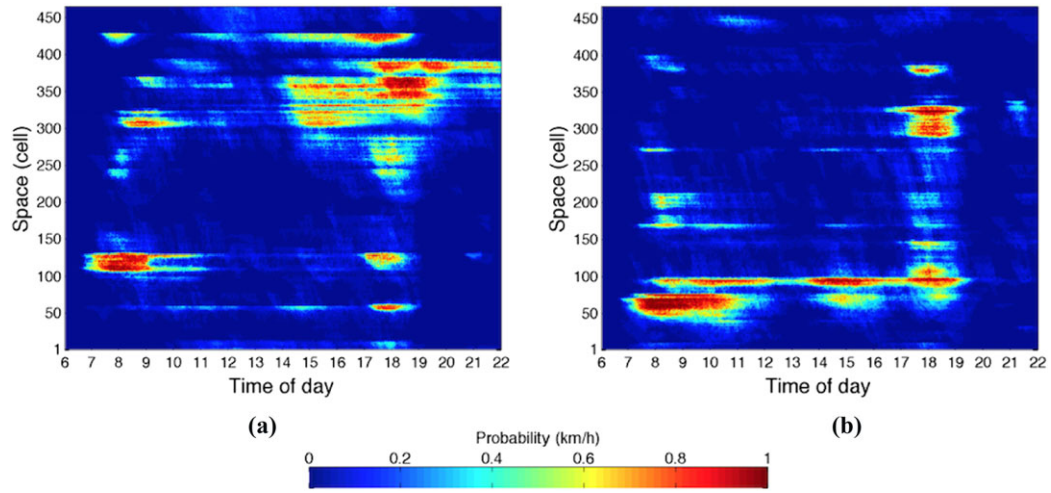

Fig. 13. Stochastic congestion maps of all 30 workdays. The spatiotemporal diagram in every single day is filtered by speed of 20 km/h, and only the clusters with more than 50 congested ST cells are reserved: (a) CW direction and (b) CCW direction.

Fig. 14. Congestion map of Beijing. The arrows indicate the major congestions, for which probability of occurrence is over 0.8. The blue arrows: morning rush hours; the red arrows: evening rush hours. The working and residential areas are roughly plotted according to the findings in Dong et al. (2015).

the recurrent congestion in Beijing. The stochastic congestion maps represent the likelihood of the occurrence of congestion based on the observations for many days (Ban et al., 2008; Yildirimoglu and Geroliminis, 2013). To construct the maps, we count the number of being in congested clusters for each ST cell during all the 30 workdays, and obtain the stochastic congestion maps through dividing the counts by the total number of the workdays; see Figure 13 for the resulting stochastic maps. In the figure, recurrent congestions can be clearly seen, and most of them occur with a really high probability that is greater than 0.8. It intuitively shows the severity of the traffic problem in Beijing. It also implies that it may not be necessary to predict the occurrence of these recurrent congestions, because they appear almost every day. Predicting the occurrence of the congestions, whose probability is about 0.5, is usually more meaningful. However, such congestions are few for ring 3 of Beijing.

Taking advantage of the stochastic congestion maps, we can stamp the detailed locations of the congestions on a real map; see Figure 14. It can be seen that the direction of most congested traffic in the morning rush hour is from the south city to the north city, and the direction in the evening rush hour is opposite. Such tides are true in Beijing, because a large number of people reside in the south side of southern ring 2 while work in its north side (Dong et al., 2015). Meanwhile, it can be observed that most of the high-probability recurrent congestions occur at the interchanges between the ring and the urban freeways connecting rings. It is not difficult to understand because the demand is usually high at the connection of two urban freeways.

4 CONCLUSION

To unveil traffic dynamics from probe vehicle data, this article proposes a simple mapping-to-cells method to construct a spatiotemporal traffic diagram for a freeway network. The method partitions a network region into

homogeneous square cells, and builds a cell network based on the number of the GPS points falling into possible downstream cells. The ranges of traffic flow directions pertaining to the cells are determined according to the moving directions of the GPS data inside cells, and the data within the range are aggregated to construct the spatiotemporal traffic diagram. After filtering by a speed threshold and clustering congested cells, major congestion and the corresponding bottlenecks can be identified. By making use of taxis' data widely existing in Beijing, the mapping-to-cells method is demonstrated and validated. In the case study, the cell network for one of the urban freeways, i.e., ring 3, is constructed, and then the traffic dynamics are visualized. Typical traffic patterns are clearly exposed such as the pinned localized cluster and oscillating congested traffic. Through the stochastic congestion maps built based on the multiday probe data, recurrent congestion and the severity of Beijing traffic are illustrated.

The advantages of the mapping-to-cells method can be summarized as follows:

- Capable. This method is capable of converting freeway networks into cell networks and extracting traffic dynamics from the scattered probe data with low temporal resolution.
- Simple. This method is simple because it is completely based on the data themselves and without the aid of any additional tool or even a digital map.
- Efficient. This method is efficient and its computational burden is low due to discretizing the continuous time and space into homogeneous square cells. It is more meaningful in the era of big data, because we usually have mass data to process.

Taking advantage of the mapping-to-cells method, we can now easily visualize traffic dynamics from probe data, which makes it possible to identify bottlenecks and predict congestion for the freeway network without installing large-scale loop detectors. Besides, partitioning a network region into cells is beneficial to incorporating multisource traffic data into the same platform, such as the data collected by bus transit and mobile phones. This will be an important direction of making use of the mapping-to-cells method in future.

However, limited by discrete nature of the square cells, the cell network cannot completely represent the continuous network in reality, which is a trade-off of the simplicity. Thus, some probe data that belong to the real network but fall out of the cell network are wasted. It results in somehow coarseness of the constructed spatiotemporal traffic diagram. However, for basically understanding the traffic, identifying bottlenecks, or even estimating travel time, the spatiotemporal diagram constructed by using the simple method might be sufficient enough. It is also noted that constructing the spatiotemporal diagram by independently using probe data is not accurate work, because the probe data are only able to convey a part of traffic information and the aggregation in space and time also results in inaccuracy.

It is worth noting that matching GPS data to a map and estimating speed along a road could be applied to extract traffic dynamics, if one has a proper digital map. Comparing with the map-matching-based methods, one of the advantages of the proposed method may still exist, i.e., efficient, because the proposed method simply maps a GPS point to a cell by using basic operations of arithmetic, while existing map-matching techniques usually have to calculate the distance of a GPS point to a line. Such advantage may make the mapping-to-cells method more powerful in the age of big data.

Extending the method on city streets with signalized intersections and quantitatively testing the constructed spatiotemporal diagram in terms of estimating travel time are all future work to enrich and better understand the mapping-to-cells method.

ACKNOWLEDGMENTS

The authors are grateful to seven anonymous referees whose comments greatly improved the quality of this article. The research is funded by NFSC (71501009, 71621001, 71501191, and 51508014).

REFERENCES

Andrienko, G. & Andrienko, N. (2008), Spatio-temporal aggregation for visual analysis of movements, in *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, Columbu, OH, 51–58.

Ban, X. J., Chu, L. & Benouar, H. (2008), Bottleneck identification and calibration for corridor management planning, *Transportation Research Record: Journal of the Transportation Research Board*, **1999**, 40–53.

Chen, C., Skabardonis, A. & Varaiya, P. (2004), Systematic identification of freeway bottlenecks, *Transportation Research Record: Journal of the Transportation Research Board*, **1867**, 46–52.

Daganzo, C. F. (1997), *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, Oxford, UK.

Dong, H., Wu, M., Ding, X., Chu, L., Jia, L., Qin, Y. & Zhou, X. (2015), Traffic zone division based on big data from mobile phone base stations, *Transportation Research Part C: Emerging Technologies*, **58**, 278–91.

Duret, A., Ahn, S. & Buisson, C. (2011), Passing rates to measure relaxation and impact of lane-changing in congestion, *Computer-Aided Civil and Infrastructure Engineering*, **26**, 285–97.

Ferreira, N., Poco, J., Vo, H. T., Freire, J. & Silva, C. T. (2013), Visual exploration of big spatio-temporal urban data: a study of New York City cab trips, *IEEE Transactions on Visualization and Computer Graphics*, **19**, 2149–58.

González, M. C., Hidalgo, C. A. & Barabási, A.-L. (2008), Understanding individual human mobility patterns, *Nature*, **453**, 779–82.

He, Z., He, S. & Guan, W. (2015a), A figure-eight hysteresis pattern in macroscopic fundamental diagrams and its microscopic causes, *Transportation Letters*, **7**, 133–42.

He, Z., Zheng, L. & Guan, W. (2015b), A simple nonparametric car-following model driven by field data, *Transportation Research Part B: Methodological*, **80**, 185–201.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q. & Bayen, A. M. (2010), Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment, *Transportation Research Part C: Emerging Technologies*, **18**, 568–83.

Hunter, T., Abbeel, P. & Bayen, A. (2014), The path inference filter: model-based low-latency map matching of probe vehicle data, *IEEE Transactions on Intelligent Transportation Systems*, **15**, 1–23.

Jenelius, E. & Koutsopoulos, H. N. (2015), Probe vehicle data sampled by time or space: consistent travel time allocation and estimation, *Transportation Research Part B: Methodological*, **71**, 120–37.

Jiang, B., Yin, J. & Zhao, S. (2009), Characterizing the human mobility pattern in a large street network, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **80**, 1–11.

Jiang, X., Zheng, C., Tian, Y. & Liang, R. (2015), Large-scale taxi O/D visual analytics for understanding metropolitan human movement patterns, *Journal of Visualization*, **18**, 185–200.

Kerner, B. S. (2009), *Introduction to Modern Traffic Flow Theory and Control*: *The Long Road to Three-Phase Traffic Theory*, Springer, Berlin, Germany.

Kerner, B. S., Rehborn, H., Aleksic, M. & Haug, A. (2004), Recognition and tracking of spatial-temporal congested traffic patterns on freeways, *Transportation Research Part C: Emerging Technologies*, **12**, 369–400.

Kerner, B. S., Rehborn, H., Schafer, R. P., Klenov, S. L., Palmer, J., Lorkowski, S. & Witte, N. (2013), Traffic dynamics in empirical probe vehicle data studied with three-phase theory: spatiotemporal reconstruction of traffic phases and generation of jam warning messages, *Physica A: Statistical Mechanics and Its Applications*, **392**, 221–51.

Laval, J., Chen, D., Amer, K. B. & Guin, A. (2009), Evolution of oscillations in congested traffic: improved estimation method and additional empirical evidences, *Transportation Research Record: Journal of the Transportation Research Board*, **2124**, 194–202.

Laval, J. & Leclercq, L. (2010), A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic, *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, **368**, 4519–41.

Li, P. & Souleyrette, R. R. (2016), A generic approach to estimate freeway traffic time using vehicle ID-matching technologies, *Computer-Aided Civil and Infrastructure Engineering*, **31**(5), 351–65.

Quddus, M. & Washington, S. (2015), Shortest path and vehicle trajectory aided map-matching for low frequency GPS data, *Transportation Research Part C: Emerging Technologies*, **55**, 328–39.

Quddus, M. A., Ochieng, W. Y. & Noland, R. B. (2007), Current map-matching algorithms for transport applications: state-of-the-art and future research directions, *Transportation Research Part C: Emerging Technologies*, **15**, 312–28.

Ramezani, M. & Geroliminis, N. (2015), Queue profile estimation in congested urban networks with probe data, *Computer-Aided Civil and Infrastructure Engineering*, **30**, 414–32.

Restrictions on Geographic Data in China (2016), https://en.wikipedia.org/wiki/Restrictions_on_geographic_data_in_China, accessed August 2016.

Tominski, C., Schumann, H., Andrienko, G. & Andrienko, N. (2012), Stacking-based visualization of trajectory attribute data, *IEEE Transactions on Visualization and Computer Graphics*, **18**, 2565–74.

Treiber, M. & Helbing, D. (2002), Reconstructing the spatio-temporal traffic dynamics from stationary detector data, *Cooperative Transportation Dynamics*, **1**, 3.1–3.4.

Treiber, M. & Kesting, A. (2013), *Traffic Flow Dynamics: Data, Models and Simulation*, Springer, Berlin, Germany.

Treiber, M., Kesting, A. & Wilson, R. E. (2011), Reconstructing the traffic state by fusion of heterogeneous data, *Computer-Aided Civil and Infrastructure Engineering*, **26**, 408–19.

Van Lint, J. & Hoogendoorn, S. P. (2010), A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways, *Computer-Aided Civil and Infrastructure Engineering*, **25**, 596–612.

Wang, Z., Lu, M., Yuan, X., Zhang, J. & Wetering, H. V. D. (2013), Visual traffic jam analysis based on trajectory data, *IEEE Transactions on Visualization and Computer Graphics*, **19**, 2159–68.

Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J., & Wu, Q. (2014), Visual exploration of sparse traffic trajectory data, *IEEE Transactions on Visualization and Computer Graphics*, **20**, 1813–22.

Wieczorek, J., Fernández-Moctezuma, R. J. & Bertini, R. L. (2009), Techniques for validating an automatic bottleneck detection tool using archived freeway sensor data, *Transportation Research Record: Journal of the Transportation Research Board*, **2160**, 87–95.

Wilson, R. E. (2008), Mechanisms for spatio-temporal pattern formation in highway traffic models, *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, **366**, 2017–32.

Yildirimoglu, M. & Geroliminis, N. (2013), Experienced travel time prediction for congested freeways, *Transportation Research Part B: Methodological*, **53**, 45–63.

Yu, L., Gao, Y., Yu, L., Song, G., Zhang, F. & Liu, J. (2015), Floating car data-based method for detecting flooding incident under grade separation bridges in Beijing, *IET Intelligent Transport Systems*, **9**, 817–23.

Zhao, N., Yu, L., Zhao, H., Guo, J. & Wen, H. (2009), Analysis of traffic flow characteristics on ring road expressways in Beijing, *Transportation Research Record: Journal of the Transportation Research Board*, **2124**, 178–85.