Analyzing the Effects of Rainfall on Urban Traffic-Congestion Bottlenecks

Yao Yao, Daiqiang Wu, Ye Hong, Dongsheng Chen, Zhaotang Liang, Qingfeng Guan, Xun Liang, and Liangyang Dai

***Abstract*—The development of geospatial big data makes it possible to study traffic-congestion issues. In particular, floating car data (FCD) is very suitable for it because FCD can help predict traffic-congestion bottlenecks and provide corresponding solutions to address traffic problems. Previous studies have discussed the impacts of rainfall on road speeds, but few studies have focused on the impacts of rainfall on the spatial distribution and changes in traffic-congestion bottlenecks throughout a mega-city. This article proposes an index calculation and clustering (ICC) model by integrating PageRank and clustering algorithms from multisource data, including rainfall data, FCD, and OpenStreetMap data. As the study area, we selected Shenzhen, which is the largest developed city in South China. The results demonstrate three peak periods of citizen travel, namely, 8:00–10:00, 14:00–16:00, and 18:00–20:00. Road speeds after rainfall decrease by 6.20% on weekdays and by 2.37% on weekends, and traffic-congestion areas increase by 23.53% and 20.65% on weekdays and on weekends, respectively. In addition, rainfall causes more significant effects on traffic conditions on weekdays compared with on weekends in Shenzhen. Compared with a traditional kernel density analysis, the proposed ICC model can offer a more thorough understanding of urban traffic-congestion areas, which can help policy makers optimize alleviation strategies.**

***Index Terms*—Clustering algorithms, floating car, geospatial big data, rainfall, traffic congestion.**

I. Introduction

With the rapid growth of urbanization in China, traffic network as the infrastructure of cities is becoming more and more important. Traffic congestion has caused many negative impacts on many cities in China, such as the overconsumption of fuel [1], increasing traffic accidents [2], [3], and air pollution [4]. A city's traffic condition is affected by various factors, including city management [5], environment conditions [6], and weather [7], [8]. Adverse weather will affect the traffic conditions not only by degrading the service level of the transportation system [9] but also by reducing the willingness of people to travel [10]. Therefore, accurately measuring the relationship between urban traffic congestion and weather is of great help in solving urban traffic problems.

Specifically, traffic flow is affected by rainfall, and vehicle speeds differ noticeably between rainy days and sunny days [11]. Some studies at different levels have shown that weather elements have more influence than other factors on public transportation [12]–[14]. For example, Smith *et al*. collected traffic and weather data and determined that traffic efficiency is related to rainfall [15]. Jia *et al.* proposed rainfall-integrated long short-term memory and the rainfall-integrated deep belief network, which could be used to model the relationship between traffic efficiency and weather, thereby predicting road speeds in different weather conditions [16]. Most studies focused on the relationship between precipitation and velocity and constructed different models to predict road speeds in different weather conditions. However, the impact of rainfall on weekends and weekdays is disparate, and the distinction remains unclear and unexplored for rainfall effects on road speeds.

Due to the availability of floating car data (FCD) in recent years [17], studying traffic congestion problems in urban transportation systems has become increasingly convenient [18]. For example, Yong-chuan *et al.* obtained road speeds with FCD to study the distribution of urban traffic congestion [19]. Li *et al.* combined FCD with cloud computing to improve the city's traffic monitoring system [20]. Some researchers have also used complex graph theory to obtain traffic status by measuring the importance of the nodes that are road intersections [21]. In a study of urban road networks, Pop *et al*. estimated the importance of nodes based on the PageRank algorithm and solved problems related to traffic congestion [22]. Based on complex networks, Yang *et al.* used the optimal weighting scheme to enhance network robustness, thus easing traffic pressure [23]. However, most studies have focused on single points and existing traffic-congestion areas, and few of them concentrates on the spatial distribution of traffic-congestion bottlenecks and have conducted analyses under different weather conditions.

This article is mainly concerned with the spatial distribution of urban transportation congestion points, areas and bottlenecks,

Manuscript received September 18, 2019; revised December 16, 2019; accepted January 11, 2020. Date of publication January 20, 2020; date of current version February 12, 2020. *(Corresponding author: Yao Yao.)* Y. Yao and D. Wu contributed equally to this work. This work was supported in part by the National Natural Science Foundation of China under Grants 41801306, 41671408, 41901332 and in part by the Natural Science Foundation of Hubei Province under Grant 2017CFA041.

Y. Yao, Q. Guan, X. Liang, and L. Dai are with the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China (e-mail: yaoy@cug.edu.cn; guanqf@cug.edu.cn; liangxun@cug.edu.cn; liangyangdai@foxmail.com).

D. Wu is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China (e-mail: daiqiang_wu@whu.edu.cn).

Y. Hong is with the Institute of Cartography and Geoinformation, ETH Zürich 8092 Zurich, Switzerland (e-mail: hongy@student.ethz.ch).

D. Chen is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: dontsingchen@whu.edu.cn).

Z. Liang is with the Institute of Space and Earth Information Science, Chinese University of Hong Kong, New Territories, Hong Kong (e-mail: liangzht3@link.cuhk.edu.hk).

Digital Object Identifier 10.1109/JSTARS.2020.2966591

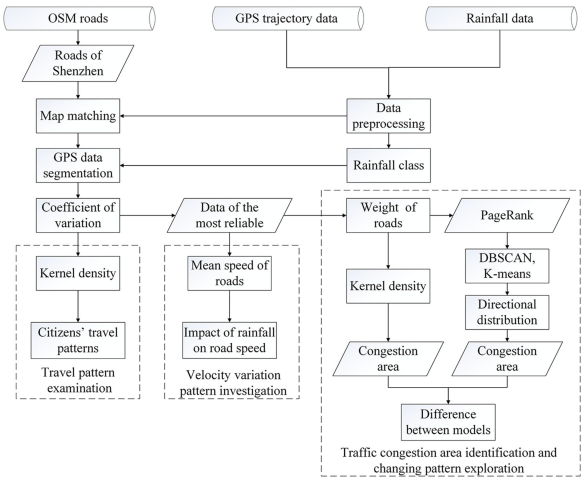

Fig. 1. Flow chart of the analysis of travel patterns and the impact of rainfall on an urban transport system.

the travel patterns of urban residents, and the impact of rainfall on traffic flows at different periods. First, citizens' aggregate travel patterns from FCD are determined. Then, we explore the impacts of rainfall on road speeds (average speed of all float cars on the road). Finally, the ICC model is proposed to examine the distribution and variation of floating cars in traffic-congestion points, areas, and bottlenecks. The findings of this article reveal the relationship between rainfall and traffic congestion and could provide valuable information for transit management.

II. Methodology

The flow chart of the proposed research design is shown in Fig. 1. This article is divided into the following three parts: 1) Examination of spatiotemporal variations in aggregate travel patterns based on kernel density and statistical analyses. After FCD OpenStreetMap (OSM) coregistration, the weekday FCD are divided into 12 groups at 2-h intervals. Then, we use a kernel density analysis to obtain the citizens' travel patterns. (2) Data period selection based on the coefficient of variation (CV) and correlation analysis between the variables based on the Pearson correlation coefficient. To determine the velocity variation pattern during rainfall, we first conduct an accuracy analysis to ensure comparability across scenarios. Each group's reliability is calculated for six sunny and six rainy days, where the groups with the highest reliability are used to compare the impacts of different weather conditions on road speeds. Moreover, we calculate the Pearson correlation coefficient to highlight the impact of rainfall on traffic conditions. (3) Identification of the traffic-congestion points and areas based on the proposed index calculation and clustering (ICC) model. The traffic-congestion points, areas, and their alterations under different scenarios are identified through the proposed ICC model. To test the effectiveness and reliability of the proposed method, a comparison with the traditional kernel density method is conducted, and the details are given in the fourth section.

A. Exploration of Citizens' Travel Patterns Based on Kernel Density Analysis

To mine travel patterns of citizens on weekdays, this article uses a kernel density analysis to conduct exploration using taxi trips. Specifically, kernel density analysis is adopted to calculate the density of features, including point kernel density (PKD) analysis and linear kernel density (LKD) analysis [24]. The PKD and LKD analyses assign the sum of the values of all the nuclear surfaces that are superimposed on a raster cell to a raster pixel by the operation of the kernel function [25]. The density of the raster cell is the sum value.

Before exploring the travel patterns of citizens in Shenzhen, FCDs that have been preprocessed are divided into 12 groups of 2-h intervals. Based on the PKD analysis, the density distribution maps of Shenzhen in different periods can be obtained. Moreover, we calculate every group's average number of GPS points. Citizens' travel patterns can be mined by analyzing the density distribution maps and numbers of GPS points in different periods.

B. Data Period Selection and Correlation Analysis

Considering the validity of the data, we need to select representative samples for the study combined with reliability indicator. Selecting representative data in research could be beneficial for workload reduction and procedure improvement. The indicator of the accuracy analysis used in this article is the CV, which is an indicator used to compare the dispersion extent of different sets of data [26]. Since the average speed differs in different periods, standard deviation will result in some incorrect judgments about the data and, thus, cannot appropriately reflect the dispersion extent of data. The CV, which is the ratio of the standard deviation to the average data speed, is a dimensionless quantity, and could eliminate this effect [27].

Meanwhile, to emphasize the impact of rainfall on traffic conditions, we adopt the Pearson correlation coefficient to analyze the relationship between different variables (including road length, road capacity, and maximum speed limitation of roads) and rainfall [28]. The Pearson correlation coefficient is an indicator used to measure the linear correlation relationship between variables [29]. The value of the Pearson correlation coefficient ranges from -1 to 1, with smaller absolute values representing less obvious correlation relationships between two variables.

In this article, the FCD for 12 days are grouped with respect to the individual time intervals. Considering the average CV of each group, the citizens' daily active period and the differences in rainfall intensity on rainy days, we select two groups whose CVs are low and for which the difference in rainfall intensity on rainy days between groups is relatively small, thereby yielding a high mean reliability. Moreover, using the correlation analysis between variables, we can highlight the impact of rainfall on the traffic conditions to enhance the reliability of the results.

C. ICC Model Based on PageRank

To further study the distribution and variation in the traffic-congestion areas and bottlenecks, this article proposes an ICC model. First, the road network is constructed to obtain the corresponding adjacency matrix, in which each intersection is transformed into a node, while a road is represented by an edge in the graph. Then, the number of floating cars is assigned as

a weight to its matched road. Finally, the weighted matrix of Shenzhen's road network is constructed.

To analyze the centrality of nodes in the graph, we adopt PageRank to sort the nodes in the graph according to their importance [30]. PageRank was initially proposed by Google and used in page ranking to sort the search results of web pages [31]. The PageRank value can be specified mathematically by the following equation [32]:

$$\boldsymbol{C}_p(v_i) = \alpha \sum_{j=1}^{n} \boldsymbol{A}_{j,i} \frac{\boldsymbol{C}_p(v_j)}{d_j^{\text{out}}} + \beta \tag{1}$$

where v_i and v_j are nodes in a graph, $\boldsymbol{C}_p(v_i)$ is the PageRank of v_i, $\boldsymbol{C}_p(v_j)$ is the PageRank of v_j, n is the in-degree of v_i, d_j^{out} is the out-degree of v_j, $\boldsymbol{A}$ is the adjacency matrix storing the structure of the graph, and α and β are fixed parameters.

The nodes with high PageRank values are considered to remain high traffic flows, which easily leads to traffic-congestion regions [33]. In this article, clustering algorithms were utilized to identify these potential regions. The process is divided into the following three steps: (1) To reduce the impacts of noise points, we utilize the DBSCAN algorithm. (2) To identify different traffic-congestion areas, we obtain various node categories through the k-means algorithm. (3) Finally, standard deviation ellipse analysis is used to identify traffic-congestion areas. Therefore, the distributions and changes in traffic-congestion areas can be obtained, which can help us predict the transfer of traffic-congestion bottlenecks.

DBSCAN is a spatial data clustering algorithm based on density [34]. While clustering, the noise nodes can be identified and removed [35]; thus, DBSCAN plays a role in eliminating some noncongested areas in the research. The k-means algorithm is a distance-based clustering algorithm that can mine the potential relationship of data [36]. At the initial stage, the cluster centers are randomly picked, and the algorithm classifies all the nodes through their dissimilarity. The center of each cluster is recalculated at each iteration, and all the nodes are reclassified based on their distance to each center. The k-means algorithm stops when the clustering results stop changing [37]. The standard deviation ellipse analysis can be used to summarize the spatial features of geographic elements, including central, discrete, and directional trends [38].

The kernel density analysis model is applied to identify the traffic-congestion areas as a comparison [39]. By identifying traffic-congestion areas and their distributions with an LKD analysis, the result of the proposed ICC model was contrasted with the result of the kernel density analysis model to evaluate its suitability for the study.

III. Study Area and Data Description

Shenzhen, China, which is located between 113°46'E to 114°37'E and 22°24'N to 22°52'N, was selected as the study area. Shenzhen is one of the four first-tier cities and one of the three national financial centers in China [40]. As shown in Fig. 2, Shenzhen has ten districts, including Futian, Luohu, Nanshan, Bao'an, Longgang, Longhua, Guangming, Yantian, Pingshan, and Dapeng [41]. Five districts, namely, Futian, Nanshan, Luohu, Bao'an, and Longgang, are the most developed areas of Shenzhen. The Longhua and Guangming Districts are newly developing areas of Shenzhen in recent years, and Yantian is a seaport district well known for its international container terminal. Historically, the eastern part of Shenzhen, including Pingshan District and Dapeng District, has been less developed due to its natural environment and limited transportation infrastructure. Specifically, according to the *Shenzhen Statistical Yearbook of 2017*, the gross domestic products of the Nanshan District, Futian District, and Longgang District are much higher than those of the other districts.

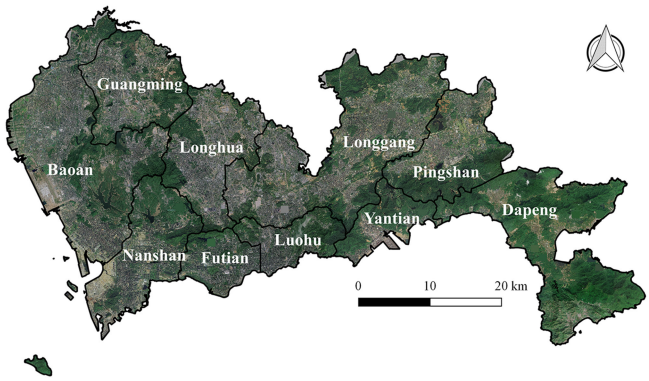

Fig. 2. Case study area: Shenzhen, China.

Due to the Chinese Economic Reform and the opening-up of the Chinese economy, Shenzhen has undergone tremendous changes [42]. As the first special economic zone established in China, it became a metropolis with significant international influence [43]. However, with the development of the economy and urbanization, the pressure of urban transport systems is increasing, resulting in negative impacts on human life [44]. Such issues pose great challenges for the functioning of urban transport systems in a Chinese metropolis [45]. An analysis and prediction of traffic-congestion points and areas can help alleviate the traffic pressures of a city.

Rainfall data are used to judge the weather conditions of Shenzhen. Rainfall data were acquired from the Shenzhen Meteorological Database (https://data.szmb.gov.cn/) and are used to facilitate day selection under different weather conditions. To derive suitable data for the study, abnormal data with large deviations from the correct locations were removed [46]. Rainfall data measured at the discrete monitoring stations were then interpolated to estimate the overall precipitation distribution of Shenzhen [47], and the results were further classified according to the standard precipitation classification metrics [48]. Based on the total precipitation summation statistics shown in Table I, we selected 12 days, namely, January 2, 3, 7, 8, 9, 11, 16, 17, 19, 28, 29, and 30, to construct a comparative analysis with different precipitation conditions [49]. The 12 days were divided into four categories: sunny weekdays (January 7, 8, and 19), rainy weekdays (January 11, 28, and 29), sunny weekends (January 2, 9, and 30), and rainy weekends (January 3, 16, and 17).

The FCD are important data in this article and were acquired from the transport commission of Shenzhen. FCD appropriately

TABLE I
RAINFALL DATA FOR JANUARY 2016 (BOLDING INDICATES SELECTED DATA)

Date		Rainfall(mm)
2016/1/1	Friday	0.007
2016/1/2	**Saturday**	**0.011**
2016/1/3	**Sunday**	**7.107**
2016/1/4	Monday	0.248
2016/1/5	Tuesday	17.299
2016/1/6	Wednesday	0.036
2016/1/7	**Thursday**	**0**
2016/1/8	**Friday**	**0.001**
2016/1/9	**Saturday**	**0.004**
2016/1/10	Sunday	0.572
2016/1/11	**Monday**	**33.634**
2016/1/12	Tuesday	0.080
2016/1/13	Wednesday	0.023
2016/1/14	Thursday	1.579
2016/1/15	Friday	30.146
2016/1/16	**Saturday**	**7.187**
2016/1/17	**Sunday**	**11.230**
2016/1/18	Monday	0.015
2016/1/19	**Tuesday**	**0**
2016/1/20	Wednesday	5.416
2016/1/21	Thursday	0.496
2016/1/22	Friday	3.414
2016/1/23	Saturday	0.355
2016/1/24	Sunday	3.784
2016/1/25	Monday	0.026
2016/1/26	Tuesday	2.156
2016/1/27	Wednesday	13.776
2016/1/28	**Thursday**	**57.553**
2016/1/29	**Friday**	**61.971**
2016/1/30	**Saturday**	**0.029**
2016/1/31	Sunday	0.079

can reflect traffic conditions well, and some scholars have utilized FCD to conduct studies about traffic conditions [40], [50]. Each FCD record contains information such as the GPS time (gpstime), plate number (plateno), longitude (lon), latitude (lat), speed, direction, effectiveness (iseffective), and the carry state (carrystate). In this article, the dates of the selected FCD are consistent with the bolded dates in Table I. This article first conducted data filtering to eliminate abnormal data, including some points from the correct locations that had large deviations. Furthermore, the effective variable can also help to remove some abnormal data (when values of iseffective are 0). Combined with a network analysis module, coregistration between FCD and OSM roads was carried out by point-to-curve and curve-to-curve matching algorithms based on the distance relationship between the points and roads and the geometric relationship between the roads and the routes of the points [51], [52]. After preprocessing the original FCD, we retrieved 210 476 308 valid data records.

Other important data for this article are the OSM road data, which are sourced from the OSM official website (http://www.openstreetmap.org). The OSM is the most abundant and

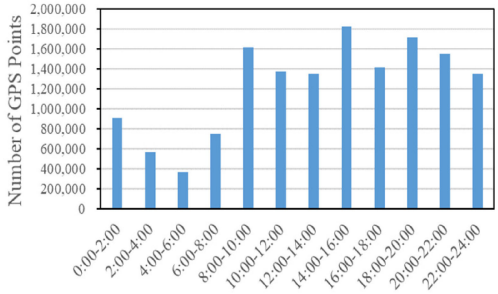

Fig. 3. Average GPS point statistics histogram on sunny weekdays.

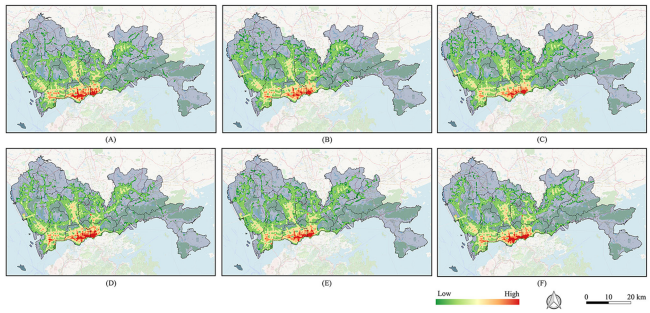

Fig. 4. FCD density distribution during 08:00–20:00 on sunny weekdays: (A) 8:00–10:00, (B) 10:00–12:00, (C) 12:00–14:00, (D) 14:00–16:00, (E) 16:00–18:00, and (F) 18:00–20:00.

effective user-generated map project generated by uploading a handheld/vehicle GPS trajectory voluntarily or through user mapping based on remote sensing images [53]. The road network data have been widely used in the field of transportation and urban functional structure research, and the related literature has demonstrated their accuracy [54]. In addition, the OSM has important attributes, such as road type, road speed, and road length, which are useful in the study of local traffic conditions [55]. Walking paths from the original OSM road data were eliminated. In addition, road data were transformed into valid junction data by satisfying topology rules (e.g., no dangles or pseudo-nodes) to construct the corresponding adjacency matrix for the proposed PageRank-based ICC model [56].

IV. RESULTS

A. *Citizens' Travel Patterns*

To obtain citizens' travel patterns on sunny weekdays, this article selected FCD from January 7, 8, and 19 of 2016. The data are classified at intervals of 2 h, and the average number of GPS points in each period is counted.

Fig. 3 reflects the number of taxi trips at different periods. From Fig. 4, we obtain citizen travel statuses in Shenzhen through user density. We can mine the following citizen travel patterns in Shenzhen on a typical weekday as follows:

1) The aggregate travel pattern of Shenzhen citizens shows strong temporal variations. Three travel peaks emerged, namely, 08:00–10:00, 14:00–16:00, and 18:00–20:00, when taxi trip numbers reached 1,619,937, 1,823,541, and

1,716,133, respectively. The numbers of taxi trips during these three periods are obviously higher than those at other times.

2) The aggregate travel pattern of Shenzhen citizens shows stable spatial hot spots. For the entire study period (8:00–20:00), the grid density of the GPS data in the Luohu District and Futian District remains consistently higher than those of other districts. For the remaining districts, the grid density of partial areas of Nanshan District, Bao'an District, Longhua District, and Longgang District (e.g., the southern region of the Bao'an District and the central region of Longhua District) is higher than those in other districts, which maintain a relatively low FCD density status.
3) An interperiod analysis could help uncover aggregated travel patterns. Our selected case shows that only the taxi trips in periods from 0:00 to 8:00 are discernibly inactive. We find that the average travel activities in the period of 18:00–24:00 are especially higher than those of the period (8:00–14:00). However, the period (14:00–18:00) remains the most highly active travel, with as many as 810 420 taxi trips on average per hour.

Here, we focused on analyzing the spatiotemporal patterns of congestion on weekdays. The results demonstrate that the travel times of the residents in Shenzhen is mainly concentrated in three periods: 08:00–10:00, 14:00–16:00, and 18:00–20:00. The number of taxi trips in Luohu District and Futian District is higher compared with those of other districts, which potentially contributes to more traffic jams and congested regions in these districts. Instead, the traffic pressure in other districts is relatively smaller. Many taxi trips continue to occur after 20:00, which could possibly indicate that people who work in Shenzhen leave work late.

TABLE II
AVERAGE COEFFICIENTS OF VARIATION (CVS) OF DIFFERENT PERIODS

Period	weekdays		weekends	
	sunny	rainy	sunny	rainy
0:00-2:00	0.466	0.443	0.448	0.448
2:00-4:00	0.525	0.504	0.473	0.475
4:00-6:00	0.587	0.538	0.537	0.528
6:00-8:00	0.525	0.494	0.483	0.488
8:00-10:00	0.559	0.515	0.484	0.471
10:00-12:00	0.507	0.513	0.514	0.494
12:00-14:00	0.485	0.501	0.513	0.507
14:00-16:00	0.521	0.546	0.546	0.565
16:00-18:00	0.549	0.572	0.547	0.565
18:00-20:00	0.579	0.621	0.563	0.542
20:00-22:00	0.497	0.522	0.523	0.514
22:00-24:00	0.455	0.472	0.475	0.472

TABLE III
QUANTITIES OF ROADS WITH DIFFERENT SPEEDS IN DIFFERENT CATEGORIES

Speed / Category	20-50 km/h	50-116 km/h
Sunny weekdays	13,741	6,214
Sunny weekends	11,999	7,585
Rainy weekdays	11,245	5,634
Rainy weekends	12,241	7,293

B. Impact of Rainfall on Road Speed

Considering the impact of other factors on traffic conditions (road length, road capacity, and maximum speed limitation of roads), we conduct a correlation analysis between different factors and rainfall. The resulting Pearson coefficients are 0.015, -0.070, and 0.058, which indicates no obvious correlation between rainfall and these factors.

After dividing the FCD into 12 groups, the average CV of each group is calculated within four categories: sunny weekdays, sunny weekends, rainy weekends, and rainy weekdays. The results in Table II show that the CV of each individual group fluctuates approximately 0.5. For the periods 0:00–2:00, 6:00–8:00, 12:00–14:00, and 22:00–24:00, the coefficients of variation are relatively low, thereby indicating that the travel reliability during these four time periods is higher. To decrease the influence of different rainfall intensities on quantization results, we conduct a statistical analysis of the average rainfall intensity of each weather station during the above four periods. The results indicate that the difference of the periods (0:00–2:00, 6:00–8:00, and 12:00–14:00) is smaller. Considering the urban lifestyle and routine resident behavior [57], this article selects the FCD during 6:00–8:00 and 12:00–14:00 to analyze the changes in road speeds and to investigate the distribution and transfer of traffic-congestion points, areas, and bottlenecks.

Table III reflects the influence of rainfall on road speeds under different weather conditions. Compared with the speed on sunny days, the overall road speed after rainfall decreased by 6.20% on weekdays and by 2.37% on weekends. The number of roads with speeds between 20 and 50 km/h after rainfall is 18.16% lower than that on sunny weekdays. However, the number of roads with the same speeds is 1.97% higher after rainfall on weekends. Compared with rainy days, on sunny days, the number of roads with speeds above 50 km/h decreases by 9.33% on weekdays and by 3.85% on weekends.

Both on weekdays and weekends, road speeds decline after rainfall. On weekdays, the change in the number of roads with certain speeds is more obvious for those roads with speeds between 20 and 50 km/h, which indicates that rainfall potentially causes greater impacts on residents in the city. Moreover, on weekends, the change in number of roads with certain speeds is more obvious for roads with speeds above 50 km/h, which indicates that rainfall could have greater effects on people who enter and leave the city.

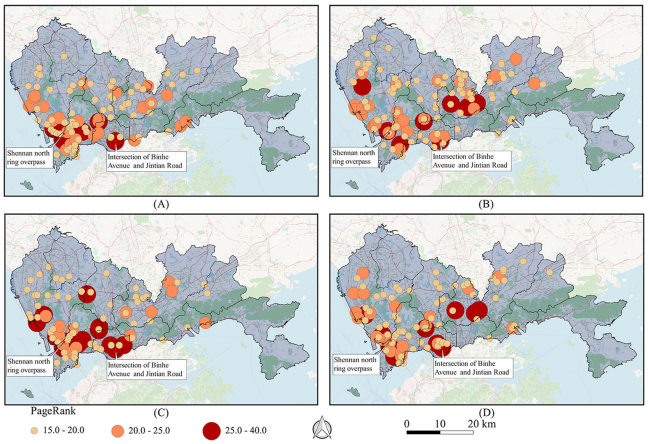

Fig. 5. Distribution of traffic-congestion points: (A) sunny weekdays, (B) rainy weekdays, (C) sunny weekends, and (D) rainy weekends.

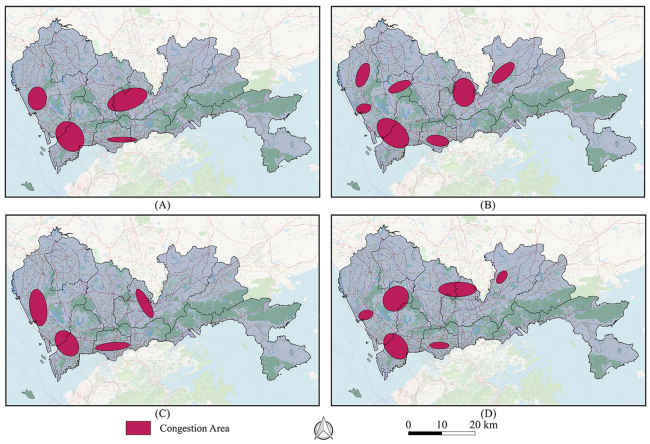

Fig. 6. Congestion distribution (ICC): (A) sunny weekdays, (B) rainy weekdays, (C) sunny weekends, and (D) rainy weekends.

TABLE IV
QUANTITIES OF CONGESTION POINTS WITH DIFFERENT LEVELS ON DIFFERENT CATEGORIES

Category \ Level	First-level	Second-level	Third-level
Sunny weekdays	75	15	7
Sunny weekends	65	10	9
Rainy weekdays	83	19	7
Rainy weekends	64	14	11

C. *Impact of Rainfall on Traffic-Congestion Condition*

Changes in traffic-congestion points, areas, and bottlenecks are analyzed by the proposed ICC model. Fig. 5 shows the quantification of three types of traffic-congestion points, which are measured by PageRank [58], on sunny and rainy days. As a data-classifying method, natural breaks could better show the differences between categories, reduce the variance within classes, and maximize the variance between classes [59]. With PageRank and natural breaks, the road network nodes were finally divided into three levels. The PageRank values of first-level points range from 15 to 20, the values of second-level points range from 20 to 25, and for third-level points, the values range from 25 to 40. More serious traffic congestion may occur at higher levels [33]. Table IV shows the count of the various level traffic-congestion points.

Whether on weekdays or on weekends, the traffic-congestion points increase after rainfall (by 12.37% and 5.95% on weekdays and weekends, respectively). In terms of the high-level PageRank points, on weekdays after rainfall, the increments of the third-level and the second-level points are 33.33%, while the increment of the first-level points is 66.66%. However, on weekends after rainfall, the number of first-level points even decreases, and the increasing numbers of traffic-congestion points are mainly second-level points. Moreover, we observe that the numbers of traffic-congestion points in the Nanshan District are far higher than those in other districts. Some traffic-congestion points are distributed at the junction of Nanshan District and Bao'an District and the junction of Futian District and Luohu District, e.g., the Shennan north ring overpass (see Fig. 5). Traffic jams also appear in the interiors of the districts, such as the intersection of Binhe Avenue and Jintian Road (see Fig. 5). In addition, Futian District, the southwestern region of Longgang District and the southern region of Bao'an District have more traffic-congestion points than the remaining districts, such as Guangming District and Yantian District. After rainfall, the traffic-congestion points of some seaside regions decrease in number, while the traffic-congestion points of midland areas increase in number.

The above results demonstrate that rainfall will have negative impacts on the traffic system. Rainfall causes a greater influence on weekdays, and the increment of low-level points is higher than that of high-level points. Traffic-congestion points are mainly distributed in the Nanshan District, which indicates that this district sustains greater traffic pressure. In terms of the variation in traffic-congestion point locations, the traffic-congestion bottlenecks transfer from the coast to the midland due to the impact of rainfall. Additionally, some urban residents would be reluctant to go out when it is rainy [60], thereby alleviating to some extent the pressure of traffic congestion, which could help explain why the low-level traffic-congestion points decrease on rainy weekends.

Traffic-congestion points will cause impacts on nearby regions. Based on the proposed ICC model framework, we use the DBSCAN and k-means clustering algorithm to obtain spatial clusters and conduct an analysis by the standard deviation ellipse to determine traffic-congestion areas. The results are shown in Fig. 6.

In Shenzhen, after rainfall, the areas of the traffic-congestion regions are 23.53% higher than that on weekdays without rainfall and are 20.65% higher than those on weekends. Traffic-congestion areas are heavily focused in four districts: Bao'an, Nanshan, Futian, and Longgang. The traffic-congestion regions in the Futian District and Nanshan District change less after rainfall. From the results, rainfall has a passive impact on the traffic

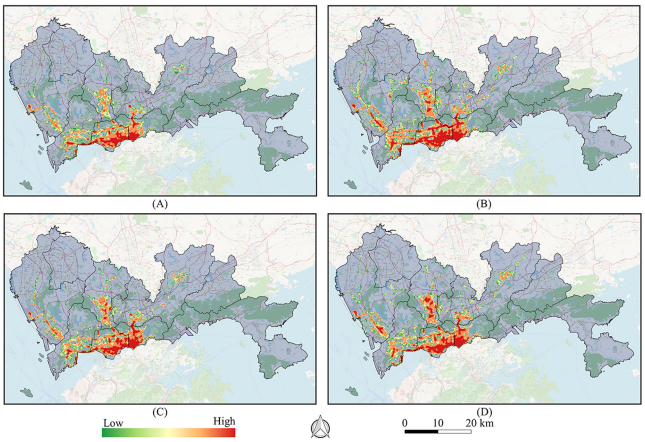

Fig. 7. Congestion distribution (LKD): (A) Sunny weekdays, (B) Rainy weekdays, (C) Sunny weekends, and (D) Rainy weekends.

system of Shenzhen. The area increment of traffic-congestion regions on weekdays is greater than that on weekends. In addition, traffic-congestion bottlenecks have a strong tendency to transfer from the coast to inland.

In comparison, a traditional kernel density analysis was implemented to evaluate the suitability of the proposed ICC model. Fig. 7 shows the distribution of traffic-congestion areas (high-density areas) through the kernel density analysis model. The area increment of traffic-congestion regions on weekdays is far greater than that on weekends. However, some traffic-congestion areas are scattered, and the change in traffic-congestion bottlenecks is not obvious. Mainly, changes after precipitation merely occur next to the existing traffic-congestion areas in a similar manner to when there is no rainfall. Compared with the proposed ICC model, which offers more specific spatial hot spots and directional changes, traditional kernel density is less informative in uncovering changes in traffic-congestion area. Therefore, the ICC model can better illustrate the regular transformational patterns of traffic conditions.

V. Discussion and Conclusion

Traffic congestion detection is significant for urban transportation science and planning applications [61]. The results demonstrate that the research framework with FCD can appropriately reflect citizens' travel patterns, and the proposed ICC model can effectively reflect the changes and distributions of all traffic-congestion points, areas, and bottlenecks.

According to the results, we find that the travel time of residents in Shenzhen is mainly concentrated over three periods (08:00–10:00, 14:00–16:00, and 18:00–20:00). During these periods, there are more taxi trips in the Futian and Luohu Districts, meaning residents of the two districts travel more frequently. The results of a correlation analysis between variables show correlations between rainfall and other factors (road length, road capacity, and maximum speed limitation of roads) are not obvious, so we highlight the impact of rainfall on traffic conditions to enhance the reliability of the results. Additionally, road speeds fall by 6.20% on weekdays and by 2.37% on weekends due to rainfall. Traffic-congestion bottlenecks transfer from the coast to the inland areas, potentially as a result of rainfall. Traffic-congestion areas also showed an increasing trend of 23.53% on weekdays and 20.65% on weekends after rainfall.

Compared with traditional research, this article contributes to developing the proposed ICC model to better determine traffic-congestion points, areas, and bottlenecks. The result of the ICC model is mainly related to the structure of the road network and the weight of the roads, which is not affected by the density of the road network. Therefore, the results of ICC model should be closer to the true traffic condition. Moreover, the model not only mines the specific traffic-congestion points, but also removes some false congestion areas caused by noise points and aggregates scattered traffic-congestion areas in a close range through clustering algorithms and a standard deviation ellipse analysis. Therefore, we can obtain more accurate and complete traffic-congestion points and areas without losing hidden traffic-congestion information. Furthermore, the ICC model highlights the trend of the overall traffic-congestion points, areas, and bottlenecks transformation and shows the distribution and change in specific traffic-congestion locations at the microlevel.

Analyzing and exploring citizens' travel patterns based on taxi trips can help us better evaluate traffic conditions [62]. Traffic congestion identification and prediction are very important for urban development, which can help optimize the allocation of resources [63]. This article mines citizens' travel patterns to obtain travel peak periods and districts, which act as a reference for the corresponding urban resource allocation. Moreover, this article contributes to proposing the ICC model to identify traffic-congestion points, areas, and bottlenecks. We can obtain specific traffic-congestion locations under different weathers through the ICC model, so effective actions can be taken in advance to alleviate the pressure of the traffic network. Furthermore, this article explored the effects of rainfall on traffic congestion, which provides a reference for traffic jam predictions during precipitation, and shows that more resources can be allocated to ease congestion in areas where rainfall is concentrated.

This article has some limitations. Currently, the results of the ICC model are obtained based on the structure of the road network and the weight of the roads. In the future, we would like to obtain urban congestion data to increase the reliability of the ICC model. Moreover, data covering longer periods will be adopted to enhance the reliability of the results. Additionally, we will compare the advantages and disadvantages of different clustering methods to obtain more accurate results. Factors such as fluctuations in traffic demand and traffic accident will affect traffic conditions to some extent, and we would like to obtain data related to traffic condition demand and traffic accident and explore the relationship between them.

Despite the above limitations, the article explores citizens' travel patterns in Shenzhen and conducts an in-depth analysis of urban traffic flows, which could offer suggestions for urban planning and recommendations for subsequent urban road network planning. With the development of ICT technology, the proposed ICC model could provide reference for real-time monitoring and prediction of traffic congestion.

REFERENCES

[1] J. I. Levy, J. J. Buonocore, and K. von Stackelberg, "Evaluation of the public health impacts of traffic congestion: a health risk assessment," *Environ. Health*, vol. 9, 2010, Art. no. 65.

[2] V. Jain, A. Sharma, and L. Subramanian, "Road traffic congestion in the developing world," in *Proc. 2nd ACM Symp. Comput. Develop.*, 2012, pp. 1–10.

[3] C. Wang, M. A. Quddus, and S. G. Ison, "Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England," *Accid. Anal. Prev.*, vol. 41, no. 4, pp. 798–808, 2009.

[4] A. T. Chin, "Containing air pollution and traffic congestion: transport policy and the environment in Singapore," *Atmos. Environ.*, vol. 30, no. 5, pp. 787–801, 1996.

[5] A. Mohan Rao and K. Ramachandra Rao, "Measuring urban traffic congestion—A review," *Int. J. Traffic Transport Eng.*, vol. 2, no. 4, pp. 286–305, 2012.

[6] T. F. Golob and W. W. Recker, "Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions," *J. Transp. Eng.*, vol. 129, no. 4, pp. 342–353, 2003.

[7] A. T. Ibrahim and F. L. Hall, "Effect of adverse weather conditions on speed-flow-occupancy relationships," *Transp. Res. Board*, 1994.

[8] M. Zhou, D. Wang, Q. Li, Y. Yue, W. Tu, and R. Cao, "Impacts of weather on public transport ridership: Results from mining data from different sources," *Transp. Res. Part C: Emerg. Technologies*, vol. 75, pp. 17–29, 2017.

[9] M. Hofmann and M. O'Mahony, "The impact of adverse weather conditions on urban bus performance measures," in *Proc. IEEE Intell. Transp. Syst.*, 2005, pp. 84–89.

[10] A. E. De Palma and D. Rochat, "Understanding individual travel decisions: Results from a commuters survey in Geneva," *Transportation*, vol. 26, no. 3, pp. 263–281, 1999.

[11] J. Zhang, G. Song, D. Gong, Y. Gao, L. Yu, and J. Guo, "Analysis of rainfall effects on road travel speed in Beijing, China," *IET Intell. Transp. Syst.*, vol. 12, no. 2, pp. 93–102, 2018.

[12] D. Akin, V. P. Sisiopiku, and A. Skabardonis, "Impacts of weather on traffic flow characteristics of urban freeways in Istanbul," *Procedia—Social Behavioral Sci.*, vol. 16, pp. 89–99, 2011.

[13] X. Li, W. H. K. Lam, M. L. Tam, and X. Cao, "Modeling the effects of rainfall intensity on traffic speed, flow, and density relationships for urban roads," *J. Transp. Eng.*, vol. 139, no. 7, pp. 758–770, 2013.

[14] W. Chung, M. Abdel-Aty, K. Choi, and J. Lee, "Application of nationwide weather data for traffic safety analysis in the United States: A spatial analysis and crash prediction modeling," in *Proc. 18th Int. Conf. Road Safety Five Continents*, Jeju Island, South Korea, May 2018, pp. 16–18.

[15] B. L. Smith, K. G. Byrne, R. B. Copperman, S. M. Hennessy, and N. J. Goodall, "An investigation into the impact of rainfall on freeway traffic flow," in *Proc. 83rd Annu. Meet. Transp. Res. Board*, Washington, DC, USA, pp. 1–13, 2004.

[16] Y. Jia, J. Wu, M. Ben-Akiva, R. Seshadri, and Y. Du, "Rainfall-integrated traffic speed prediction using deep learning method," *Iet Intell. Transp. Syst.*, vol. 11, no. 9, pp. 531–536, 2017.

[17] Y. Huang, D. J. Sun, and L. Zhang, "Effects of congestion on drivers' speed choice: Assessing the mediating role of state aggressiveness based on taxi floating car data," *Accident Anal. Prevention*, vol. 117, pp. 318–327, 2018.

[18] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Physica A: Statist. Mech. Its Appl.*, vol. 477, pp. 149–160, 2017.

[19] Z. Yong-chuan, Z. Xiao-qing, Z. Li-ting, and C. Zhen-ting, "Traffic congestion detection based on GPS floating-car data," *Procedia Eng.*, vol. 15, pp. 5541–5546, 2011.

[20] Q. Li, T. Zhang, and Y. Yu, "Using cloud computing to process intensive floating car data for urban traffic surveillance," *Int. J. Geographical Inf. Sci.*, vol. 25, no. 8, pp. 1303–1322, 2011.

[21] S. P. Borgatti, "Centrality and network flow," *Social Netw.*, vol. 27, no. 1, pp. 55–71, 2005.

[22] F. Pop and C. Dobre, "An efficient pagerank approach for urban traffic optimization," *Math. Probl. Eng.*, vol. 2012, 2012, Art. no. 465613.

[23] R. Yang, W. Wang, Y. Lai, and G. Chen, "Optimal weighting scheme for suppressing cascades and traffic congestion in complex networks," *Physical Rev. E, Statist., Nonlinear, Soft Matter Phys.*, vol. 79, no. 2 Pt 2, 2009, Art. no. 26112.

[24] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Abingdon, U.K.: Routledge, 2018.

[25] D. E. Seaman and R. A. Powell, "An evaluation of the accuracy of kernel density estimators for home range analysis," *Ecology*, vol. 77, no. 7, pp. 2075–2085, 1996.

[26] R. van Zyl and A. J. van der Merwe, "A Bayesian control chart for a common coefficient of variation," *Commun. Stat.-Theor. M.*, vol. 46, no. 12, pp. 5795–5811, 2017.

[27] A. Yan, S. Liu, and M. Azam, "Designing a multiple state repetitive group sampling plan based on the coefficient of variation," *Commun. Statist.-Simul. Comput.*, vol. 46, no. 9, pp. 7154–7165, 2017.

[28] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. New York, NY, USA: Springer, 2009, pp. 1–4.

[29] J. Adler and I. Parmryd, "Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient," *Cytom. Part A*, vol. 77, no. 8, pp. 733–742, 2010.

[30] X. Meng, Z. Li, and Z. Yang, "BRING PageRank TO the infrastructure network," *Acta Phys. Pol. B.*, vol. 49, no. 7, pp. 1497–1505, 2018.

[31] S. Kamvar, T. Haveliwala, and G. Golub, "Adaptive methods for the computation of PageRank," *Linear Algebra Appl.*, vol. 386, pp. 51–65, 2004.

[32] M. Kale and P. S. Thilagam, "DYNA-RANK: Efficient Calculation and Updation of PageRank," in *Proc. Int. Conf. Comput. Sci. Inf. Technol.*, 2008, pp. 808–812.

[33] Y. Yao, Y. Hong, D. Wu, Y. Zhang, and Q. Guan, "Estimating the effects of "community opening" policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses," *Comput., Environ. Urban Syst.*, vol. 70, pp. 163–174, 2018.

[34] A. Joshi and R. Kaur, "A review: Comparative study of various clustering techniques in data mining," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 3, pp. 55–57, 2013.

[35] Q. Rong, J. Yan, and G. Guo, "Research and implementation of clustering algorithm based on DBSCAN," *Comput. Appl.*, vol. 4, pp. 45–46, 2004.

[36] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[37] K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," *Elect. Eng. Comput. Sci.*, 1997.

[38] D. W. Lefever, "Measuring geographic concentration by means of the standard deviational ellipse," *Am. J. Sociology*, vol. 32, no. 1, pp. 88–94, 1926.

[39] A. Tabibiazar and O. Basir, "Kernel-based optimization for traffic density estimation in ITS," in *Proc. IEEE Veh. Technol. Conf.*, 2011, pp. 1–5.

[40] D. R. Meyer, "Shenzhen in China's financial center networks," *Growth Change*, vol. 47, no. 4, pp. 572–595, 2016.

[41] G. Cheng *et al.*, "Spatial difference analysis for accessibility to high level hospitals based on travel time in Shenzhen, China," *Habitat Int.*, vol. 53, pp. 485–494, 2016.

[42] D. Wall, "China's economic reform and opening-up process: The role of the special economic zones," *Develop. Policy Rev.*, vol. 11, no. 3, pp. 243–260, 1993.

[43] F. Chuang-Lin, "The urbanization and urban development in China after the reform and opening-up," *Econ. Geography*, vol. 29, no. 1, pp. 19–25, 2009.

[44] W. Xiang, Y. Wang, N. Li, and Q. Zhu, "Grey-relation analysis of traffic system and urbanization in Jilin Province of China," *Chin. Geogr. Sci.*, vol. 17, no. 3, pp. 216–221, 2007.

[45] J. Wang and D. He, "Sustainable urban development in China: Challenges and achievements," *Mitig. Adapt. Strat. Gl.*, vol. 20, no. 5, pp. 665–682, 2015.

[46] M. F. Hutchinson, "Interpolation of rainfall data with thin plate smoothing splines. Part I: Two dimensional smoothing of data with short range correlation," *J. Geographic Inf. Decis. Anal.*, vol. 2, no. 2, pp. 139–151, 1998.

[47] K. N. Dirks, J. E. Hay, C. D. Stow, and D. Harris, "High-resolution studies of rainfall on Norfolk Island: Part II: Interpolation of rainfall data," *J. Hydrol.*, vol. 208, nos. 3/4, pp. 187–193, 1998.

[48] K. Bartoszek and D. Skiba, "Circulation types classification for hourly precipitation events in Lublin (East Poland)," *Open Geosci.*, vol. 8, no. 1, pp. 214–230, 2016.

[49] Q. Li, X. Hao, W. Wang, A. Wu, and Z. Xie, "Effects of the rainstorm on urban road traffic speed – A case study of Shenzhen, China," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 71–75, 2017.

[50] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Gener. Comput. Syst.*, vol. 61, pp. 97–107, 2016.
[51] A. Zhang, Z. Gao, and H. Ren, "Incident-based traffic congestion control strategy," *Sci. China Technological Sci.*, vol. 54, no. 5, pp. 1338–1344, 2011.
[52] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Integrity of map-matching algorithms," *Transp. Res. Part C: Emerg. Technologies*, vol. 14, no. 4, pp. 283–302, 2006.
[53] J. J. Arsanjani, A. Zipf, P. Mooney, and M. Helbich, "An introduction to OpenStreetMap in geographic information science: Experiences, research, and applications," *OpenStreetMap in GIScience*. New York, NY, USA: Springer, 2015, pp. 1–15.
[54] J. Mondzech and M. Sester, "Quality analysis of OpenStreetMap data based on application needs," *Cartographica: Int. J. Geographic Inf. Geovisualization*, vol. 46, no. 2, pp. 115–125, 2011.
[55] D. Zielstra and A. Zipf, "Quantitative studies on the data quality of OpenStreetMap in Germany." *Proceedings of GIScience*. vol. 2010, no. 3. 2010.
[56] A. Sevtsuk and M. Mekonnen, "Urban network analysis: A new toolbox for measuring city form in ArcGIS," in *Proc. Symp. Simul. Archit. Urban Design*, 2012, pp. 1–10.
[57] J. Xiao, H. Li, X. Wang, and S. Yuan, "Traffic peak period detection from an image processing view," *J. Adv. Transp.*, vol. 2018, pp. 1–9, 2018.
[58] T. Agryzkov, J. L. Oliver, L. Tortosa, and J. F. Vicent, "An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector," *Appl. Math. Comput.*, vol. 219, no. 4, pp. 2186–2193, 2012.
[59] J. Chen, S. Yang, H. Li, B. Zhang, and J. Lv, "Research on geographical environment unit division based on the method of natural breaks (Jenks)," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XL-4/W3, pp. 47–50, 2013.
[60] A. Lockwood, S. Srinivasan, and C. Bhat, "Exploratory analysis of weekend activity patterns in the San Francisco Bay Area, California," *Transp. Res. Rec.: J. Transp. Res. Board*, vol. 1926, pp. 70–78, 2005.
[61] K. Mandal, A. Sen, A. Chakraborty, S. Roy, S. Batabyal, and S. Bandyopadhyay, "Road traffic congestion monitoring and measurement using active RFID and GSM technology,"in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst.*, 2011, pp. 1375–1379.
[62] J. Tang, F. Liu, Y. Wang, and H. Wang, "Uncovering urban human mobility from large scale taxi GPS data," *Physica A: Statist. Mech. Appl.*, vol. 438, pp. 140–153, 2015.
[63] R. Bauza, J. Goz, and A. Lvez, "Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications," *J. Netw. Comput. Appl.*, vol. 36, no. 5, pp. 1295–1307, 2013.

Yao Yao received the B.S. degree in surveying and mapping engineering and the M.S. degree in geodesy and surveying engineering from Wuhan University, Wuhan, China, in 2008 and 2011, respectively. He received the Ph.D. degree in cartography and geographic information system from Sun Yat-sen University, Guangzhou, China, in 2017.

He is currently an Associate Professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China. His research interests include the application of geospatial big data and urban computing.

Daiqiang Wu received the B.S. degree in surveying and mapping engineering from Sun Yat-sen University, Guangzhou, China, in 2018. He is currently working toward the M.S. degree in cartography and geographic information system with School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

His current research interests include scene classification and remote sensing application.

Ye Hong received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2018. He is currently working toward the M.S. degree with the Institute of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland.

His research interests include machine learning and deep learning in geospatial big data mining and intelligent understanding of remote sensing images.

Dongsheng Chen received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2018. He is currently working toward the M.S. degree in cartography and geographic information system with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University.

His research interests include machine learning, big data mining, and socio-geographic calculation.

Zhaotang Liang received the B.S. degree in surveying and mapping engineering from Sun Yat-sen University, Guangzhou, China, in 2017 and the M.S. degree from the Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, in 2018.

His research interests include the application of geospatial big data and urban computing.

Qingfeng Guan received the B.S. degree in geography from East China Normal University, Shanghai, China, in 2000, the M.S. degree in geographic information science and cartography from Chinese Academy of Sciences, Beijing, China, in 2003, and the Ph.D. degree in geographic information science and cartography from University of California, Santa Barbara, CA, USA, in 2008.

He is currently a full Professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China. His research interests include the application of geospatial big data and high-performance spatial intelligence computation.

Xun Liang received the M.S and Ph.D. degrees in cartography and geographic information system from Sun Yat-sen University, Guangzhou, China, in 2015 and 2018.

His research interests include the future land-use simulation and machine learning.

Liangyang Dai is currently working toward the graduate degree (major in geographic information science) from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China.