# ANALYZING THE EFFECTS OF RAINFALL ON THE URBAN TRAFFIC CONGESTION BOTTLENECKS BY USING FLOATING CAR DATA

*Yao Yao[1,3,*], Daiqiang Wu[2], Jian Yang[1], Xuguo Shi[1], Lin Du[1], Yuyang Cai[1]*

1. School of Geography and Information Engineering, China University of Geoscience, Wuhan 430074.
2. School of Remote Sensing and Information Engineering, Wuhan University. Wuhan 430075.
3. Alibaba Group, Hangzhou 311121.

## ABSTRACT

The development of geospatial big data makes it possible to study traffic congestion issues through floating car data (FCD). FCD can help predict the traffic congestion bottlenecks and provide corresponding solutions to address traffic problems. However, few studies have focused on the distribution and changes in traffic congestion bottlenecks throughout a mega-city. This study proposes an index calculation and clustering (ICC) model by integrating PageRank and clustering algorithms via multisource data, including rainfall data, FCD and OpenStreetMap data. We selected Shenzhen, the largest developed city in South China, as the study area. The results demonstrate that there are three peak periods of citizen travel: 8:00-10:00, 14:00-16:00, and 18:00-20:00. Road speeds after rainfall decrease, and traffic congestion areas increase. The results also quantitatively analyzed the differences of traffic congestion between work day and rest day. The proposed ICC model can offer a thorough understanding of urban traffic congestion areas, which can help policy makers optimize alleviation strategies.

***Index Terms—*** Geospatial big data, traffic congestion, rainfall, floating car data

## 1. INTRODUCTION

With the rapid growth of Chinese urbanization, traffic congestion has caused many negative impacts in Chinese cities, such as overconsumption of fuel[1], increases in traffic accidents[2], and air pollution[3]. To our knowledge, a city's traffic condition is affected by various factors, including city management, population density, and weather[4]. Accurately measuring the relationship between urban traffic congestion and environmental factors is of great help in solving urban traffic problems.

Specifically, traffic flow is affected by rainfall. Some studies have shown that traffic efficiency and weather are firmly relevant[5]. For example, Smith et al. (2004) collected traffic and weather data and determined that traffic efficiency is related to rainfall[6]. Stamos et al. (2016) demonstrated that rainfall could cause significant reductions in car speed[7]. Most studies focused on the relationship between precipitation and velocity in different weather conditions. However, the impact of rainfall on rest days and work days is disparate. Such a distinction between work days and rest days remains unclear and unexplored for rainfall effects on road speed.

In recent years, based on FCD, studying traffic congestion problems in urban transportation systems with complex graph theory has become increasingly convenient[8]. For example, Yong-chuan et al. (2011) obtained the average speed of roads with FCD to study the distribution of urban traffic congestion on the basis of point-to-curve and curve-to-curve map matching algorithms[9]. Li et al. (2011) combined FCD with cloud computing to improve the city's traffic monitoring system[10]. In a study of urban road networks, Pop et al. (2012) estimated the importance of nodes based on the PageRank algorithm and solved problems related to traffic congestion[11]. Based on complex networks, Yang et al. (2009) used the optimal weighting scheme to enhance network robustness, thus easing traffic pressure[12]. However, most of the studies have focused on single points instead of traffic congestion areas, and few of them conducted analyses under different weather conditions and considered work day/rest day effects.

To explore relationship between rainfall and traffic condition, this study combines FCD, OpenStreetMap road data and rainfall data with ICC model and some statistics methods to summarize citizens' travel patterns, and analyze the impact of rainfall on traffic flow at different time periods, identify existing and potential traffic congestion areas.

## 2. DATA DESCRIPTION AND METHODOLOGY

### 2.1. Data description

As one of the four first-tier cities and one of the three national financial centers in China, Shenzhen, China, was selected as the study area. As shown in Figure 1, Shenzhen has ten districts, including Futian, Luohu, Nanshan, Yantian and so on[13].

This paper used multi-source data, including floating car data, rainfall data and OpenStreetMap (OSM) road data, to analyze citizens' patterns and the impact of rainfall on traffic flow at different time periods. FCD is one of the most important pieces of data in this study and was acquired from the transport commission of Shenzhen. Rainfall data are used to judge the weather conditions of Shenzhen in this study and were acquired from the Shenzhen Meteorological Database (https://data.szmb.gov.cn/). OSM road data are sourced from the OSM official website (http://www.openstreetmap.org). OpenStreetMap road data is the most abundant and effective road data, and have been widely used in the field of transportation and urban functional structure research.

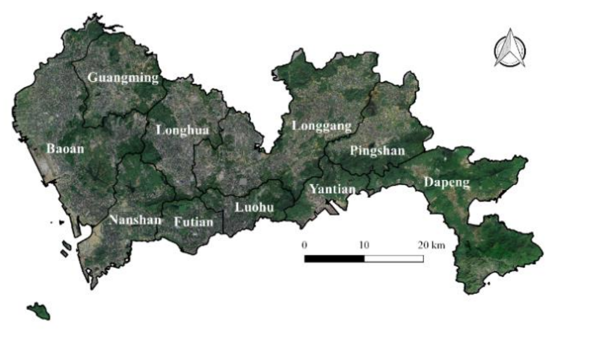


Figure 1 Case study area: Shenzhen, China.

## 2.2. Methodology

The study is divided into the following three parts: (a) Examination of spatiotemporal variations in aggregate travel patterns based on kernel density and statistical analysis. (b) Data period selection based on the coefficient of variation. (c) Traffic congestion area identification based on the proposed index calculation and clustering (ICC) model.

Before exploring the travel patterns of citizens in Shenzhen, FCD needs to be preprocessed and map matched with OSM road data to remove abnormal data. Then, the FCD is divided into twelve groups at intervals of two hours. On the basis of point kernel density analysis, the density distribution maps of Shenzhen in the different time periods could be obtained. At the same time, we calculate every group's number of GPS points. Finally, we can mine citizens' travel patterns.

Considering the validity of the data, we need to select representative samples for the study combined with indicators of reliability. The indicator of reliability analysis used in this paper is the coefficient of variation (CV)[14]. In this study, the FCD of four days is grouped with respect to each individual time interval. The CV of each group is calculated. Considering the CV of every group and daily life period, the data of four days are comprehensively analyzed to select two groups whose CVs are relatively low, which means high reliability. The second and third parts of the study obtain more accurate results with the two groups of data.

To further study the distribution and variation in traffic congestion areas and bottlenecks, this study proposed an effective index calculation and clustering (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 transformed into edge in a graph. Then we calculated PageRank of each node, and the nodes with high PageRank values are considered to remain as high traffic flow, which easily leads to traffic congestion regions[15]. Next we removed some noise nodes by using DBSCAN, and obtain classified nodes with different categories through K-means algorithm. Finally, we use standard deviation ellipse analysis to identify traffic congestion areas.

# 3. RESULTS

## 3.1. Citizens' travel patterns

The data on January 7, 2016 are classified at intervals of two hours, and the number of GPS points in each time period is counted. The results are shown in Figure 2.

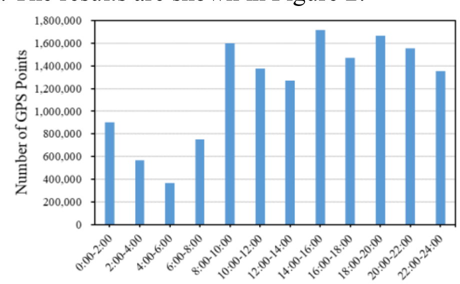


Figure 2 GPS point statistics histogram on January 7, 2016

Combining the histogram in Figure 2 with the FCD density distribution in Figure 3, we can mine citizens' travel patterns in Shenzhen on a typical work day. Three travel peaks emerged at periods of 08:00-10:00, 14:00-16:00 and 18:00-20:00, when the taxi trip numbers reached 1,599,965, 1,713,640 and 1,667,182, respectively. Our selected case shows that only taxi trips in periods from 0:00 to 8:00 are discernibly inactive. In addition, we find that the average travel activities in the period of 18:00-24:00 are particularly higher than those of the period of 8:00-14:00. And the period (14:00-18:00) remains the most highly active travel, with as much as 795,690 taxi trips on average per hour. And the aggregate travel pattern of Shenzhen citizens shows stable spatial hot spots. For the entire study time period (8:00 - 20:00), the grid density of GPS data in the Luohu District and Futian District remains consistently higher than those of other districts.

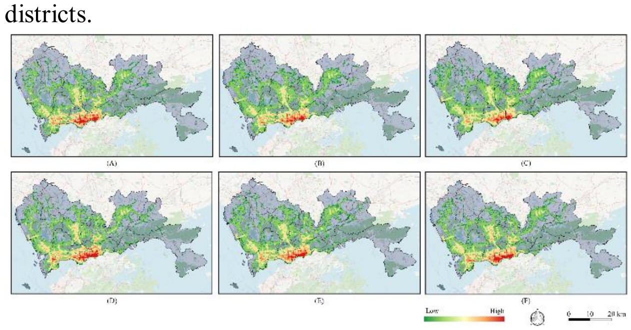

Figure 3 FCD density distribution during 08:00-20:00 on January 7: (A) 8:00-10:00, (B) 10:00-12:00, (C) 12:00-14:00, (D) 14:00-16:00, (E) 16:00-18:00, and (F) 18:00-20:00.

### 3.2. Impact of rainfall on road speed

After dividing the FCD into 12 groups, the coefficient of variation (CV) of each group is calculated within four days. Considering the urban lifestyle and resident routine behavior[16], this study finally selects FCD during 6:00-8:00 and 12:00-14:00 to analyze the changes in road speeds and investigate the distribution and transfer of traffic jam areas and bottlenecks.

Figure 4 reflects the influence of rainfall on road speed under different weather conditions. Compared with sunny day, after rainfall, the overall road speed decreased by 6.32% on work days and 2.42% on rest days. In addition, we found that the road network densities in Nanshan District, Futian District and Luohu District are higher compared with those in other districts. The roads of Futian District and Luohu District are mainly low speed. Overall, the roads are concentrated in the central and southwestern regions of Shenzhen.

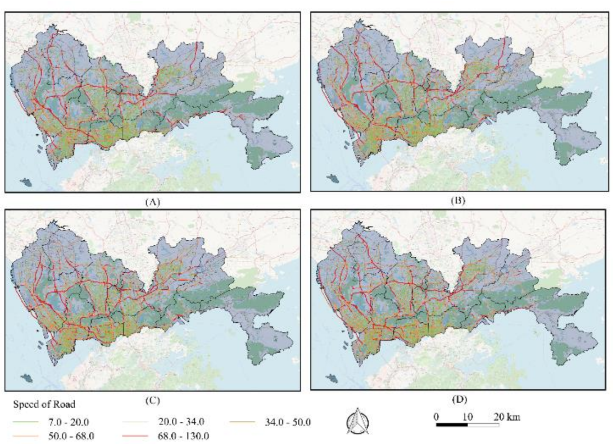


Figure 4 Speed of road network (km/h): (A) January 7 (sunny and work day), (B) January 28 (rainy and work day), (C) January 9 (sunny and rest day), and (D) January 17 (rainy and rest day).

### 3.3. Impact of rainfall on traffic congestion area

The changes in traffic congestion areas and bottlenecks are analyzed by the proposed ICC model. Figure 5 shows the quantification of three types of traffic congestion points, which are measured by PageRank[17], on sunny days and on rainy days. Some traffic congestion points are distributed in the junction of the Nanshan District and Bao'an District and the junction of Futian District and Luohu District, such as the Shennan north ring overpass (Figure 5). Additionally, some traffic jams also appear in the interior of some districts, such as the intersection (Figure 5) of Binhe Avenue and Jintian Road. In addition, Futian District, the southwestern region of Longgang District and the southern region of Bao'an District have more traffic congestion points. After rainfall, traffic congestion points of some seaside regions decrease, while traffic congestion points of midland increase.

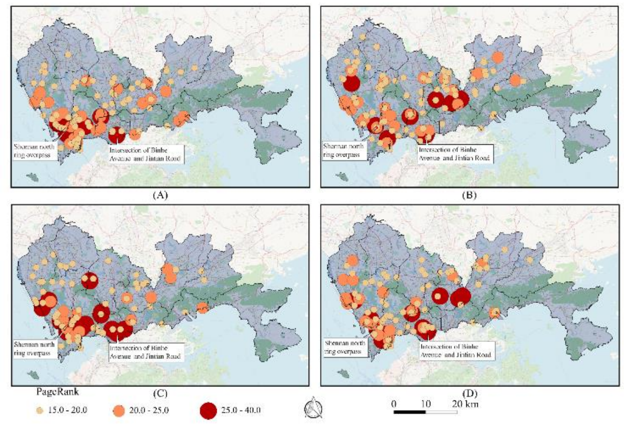


Figure 5 Distribution of traffic congestion points: (A) January 7 (sunny and work day), (B) January 28 (rainy and work day), (C) January 9 (sunny and rest day), and (D) January 17 (rainy and rest day).

Traffic congestion points will cause impacts on nearby regions. Based on the proposed ICC model framework, we determined traffic congestion areas. The results are shown in Figure 6. After rainfall, the area of the traffic congestion regions is 23.0% higher than that on work days without rainfall and is 20.70% higher than that on rest days without rainfall. Traffic congestion areas are heavily focused in four districts: Bao'an, Nanshan, Futian and Longgang Districts. For the overall trend, traffic congestion bottlenecks have a strong tendency to transfer from the coast to inland. And some of the larger traffic congestion areas are broken into several small congestion areas due to the impacts of rainfall.

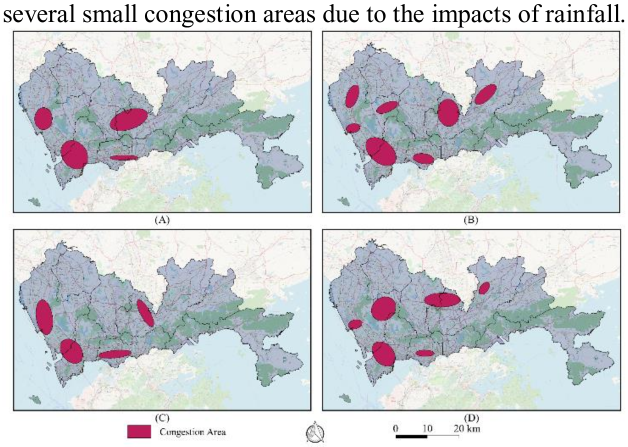


Figure 6 Congestion distribution: (A) January 7 (sunny and work day), (B) January 28 (rainy and work day), (C) January 9 (sunny and rest day), and (D) January 17 (rainy and rest day).

## 4. CONCLUSION

Traffic congestion detection is significant for urban transportation science and planning applications[18]. This study utilizes FCD, rainfall data and OSM road data to obtain citizens' travel patterns in Shenzhen and explore the impact of rainfall on both urban road speeds and traffic congestion areas. The results demonstrate that the research framework with FCD can reflect citizens' travel patterns well, and the proposed ICC model can effectively reflect the changes and distribution of traffic congestion areas and bottlenecks. Based on the analysis results on the alteration of traffic congestion areas, this study provides a feasible solution to predict potential traffic congestion bottlenecks.

According to the results, we find that the travel time of residents in Shenzhen is mainly concentrated over three time periods (08:00-10:00, 14:00-16:00 and 18:00-20:00) and taxi trips in Futian District and Luohu District are more likely. So at the peak periods of people's travel, the investment in transportation can be appropriately increased. In some districts where traffic congestion is likely to occur, public transportation should be vigorously developed to decrease traffic congestion. At the same time, the results indicate that rainfall has a great impact on urban traffic conditions. Combined with changes of traffic congestion areas, we should take effective actions in advance to alleviate the pressure of traffic congestion. For example, we could improve the road network structure, effectively manage non-motor vehicles, and increase fuel tax prices according to specific city traffic conditions to improve traffic congestion.

The study explores citizens' travel patterns in Shenzhen and conducts an in-depth analysis of urban traffic flows, which could provide suggestions for urban planning and recommendations for subsequent urban road network planning. In the future, the proposed ICC model could serve as a reference for urban traffic flow analysis in the field of transportation geography.

## 5. REFERENCES


[1] Levy, J.I., Buonocore, J.J. and von Stackelberg, K., 2010. Evaluation of the public health impacts of traffic congestion: a health risk assessment. *Environ Health*, 9, 65.

[2] Wang, C., Quddus, M.A. and Ison, S.G., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accid Anal Prev*, 41(4), 798-808.

[3] Chin, A.T., 1996. Containing air pollution and traffic congestion: transport policy and the environment in Singapore. *Atmospheric Environment*, 30(5), 787-801.

[4] Mohan Rao, A. and Ramachandra Rao, K., 2012. Measuring urban traffic congestion-a review. *International Journal for Traffic and Transport Engineering*, 2(4), 286-305.

[5] Li, X., et al., 2013. Modeling the Effects of Rainfall Intensity on Traffic Speed, Flow, and Density Relationships for Urban Roads. *Journal of Transportation Engineering*, 139(7), 758-770.

[6] Smith, B.L., et al., 2004. An investigation into the impact of rainfall on freeway traffic flow. *83rd Annual Meeting of the Transportation Research Board, Washington, DC.*

[7] Stamos, I., et al., 2016. Modeling Effects of Precipitation on Vehicle Speed. *Transportation Research Record: Journal of the Transportation Research Board*, 2551, 100-110.

[8] Yan, Y., et al., 2017. Understanding characteristics in multivariate traffic flow time series from complex network structure. *Physica A: Statistical Mechanics and its Applications*, 477, 149-160.

[9] Yong-chuan, Z., et al., 2011. Traffic Congestion Detection Based On GPS Floating-Car Data. *Procedia Engineering*, 15, 5541-5546.

[10] Li, Q., Zhang, T. and Yu, Y., 2011. Using cloud computing to process intensive floating car data for urban traffic surveillance. *International Journal of Geographical Information Science*, 25(8), 1303-1322.

[11] Pop, F. and Dobre, C., 2012. An efficient pagerank approach for urban traffic optimization. *Mathematical Problems in Engineering*, 2012.

[12] Yang, R., et al., 2009. Optimal weighting scheme for suppressing cascades and traffic congestion in complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 79(2 Pt 2), 26112.

[13] Cheng, G., et al., 2016. Spatial difference analysis for accessibility to high level hospitals based on travel time in Shenzhen, China. *Habitat International*, 53, 485-494.

[14] Pu, W., 2011. Analytic relationships between travel time reliability measures. *Transportation Research Record: Journal of the Transportation Research Board*(2254), 122-130.

[15] Yao, Y., et al., 2018. Estimating the effects of "community opening" policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses. *Computers, Environment and Urban Systems*, 70, 163-174.

[16] Xiao, J., et al., 2018. Traffic Peak Period Detection from an Image Processing View. *Journal of Advanced Transportation*, 2018, 1-9

[17] Agryzkov, T., et al., 2012. An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector. *Applied Mathematics and Computation*, 219(4), 2186-2193.

[18] Mandal, K., et al., 2011. Road traffic congestion monitoring and measurement using active RFID and GSM technology. *Intelligent Transportation Systems (ITSC)*, 1375-1379.