

## HOW AIR QUALITY AFFECTS HUMAN MOBILITY PATTERNS: AN EXPLORATORY ANALYSIS

Shishuo Xu <sup>1,2</sup>, Yang Zhao <sup>1,2,\*</sup>, Songnian Li <sup>3</sup>

<sup>1</sup> School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 15 Yongyuan Road, Beijing, 102616, China; xushishuo@bucea.edu.cn; 2108160322007@stu.bucea.edu.cn

<sup>2</sup> Key Laboratory of Urban Spatial Informatics, Ministry of Natural Resources of the People's Republic of China, 15 Yongyuan Road, Beijing, 102616, China

<sup>3</sup> Department of Civil Engineering, Toronto Metropolitan University, 350 Victoria Street, Toronto, M5B 2K3 Canada; snli@torontomu.ca

**KEY WORDS:** human mobility, air quality, spatiotemporal analysis, correlation analysis, bike sharing data.

### ABSTRACT:

Air quality acts as an important factor that human may consider as they make decisions on when and where they would go. In order to access how much the air quality affects human mobility patterns, the air quality was measured using air quality index (AQI) and human mobility patterns were measured by travel volume and travel distance of shared bikes. Their correlation that presents on weekdays and weekends as well as in different administrative districts were investigated using Spearman correlation analysis method. A case study was conducted in Beijing, China using bike sharing data and air quality data ranging from May 10 to 16, 2017. The results show that travel distance is more sensitive to air quality on weekdays such as Changping District (-0.20), Haidian District (-0.13), Shunyi District (-0.12). The travel volume on weekdays is less sensitive to air quality due to commuting. The travel volume has a negative relationship with AQI on weekends. Fengtai District, Huairou District, Pinggu District are more susceptible to severe air quality, leading to a reduction in bike traveling distance. This work sheds light on understanding human-environment coupling mechanism and promoting urban sustainable development.

### 1. INTRODUCTION

Understanding human mobility patterns plays a key role in a wide range of fields, such as urban planning (Kang et al., 2012; Liu et al., 2012), traffic forecasting (Peng et al., 2012), disease spreading (Bagrow et al., 2012), location-based recommendation systems (Cheng et al., 2011), and so on. Air quality is a concern for all countries, as severe air pollution has adverse effects on human health. It is reported that only one percent of 500 cities in China satisfy air quality standard set by the World Health Organization (WHO) (Peng et al., 2014). Some students were absent from school and workers stopped working due to the deteriorating air conditions (Ma et al., 2019). To minimize the risks of diseases caused by severe air pollution, people are likely to consider air quality when engaging in outdoor activities. For example, if there is no hazardous alert for a week or more days, the good air condition has little impact on human movement. In contrast, if the bad air pollution frequently happens, people are more likely to participate in indoor activities instead of going outside for exercise or entertainment. Therefore, air quality can be considered as one of the factors influencing human mobility patterns, especially in cities with severe air pollution.

Some existing studies explore the influence of external factors (e.g., weather) on human mobility patterns using logistic regression models (e.g., Logit model) based on the questionnaire data, which is time consuming and inefficient (Campbell et al., 2016). With the development of mobile phone and Internet technologies, a number of existing studies have investigated the impact of air quality on human mobility patterns using social media check-ins (Yan et al., 2019), social sensor data (Sagl et al., 2012), public transportation smart card data (Ma et al., 2017), subway passenger data (Wu et al., 2020), taxi data (Kang et al., 2019), mobile positioning records (Xu et al., 2021) and docked bike sharing data (Cao et al., 2019; Gebhart et al., 2014), where dockless bike sharing data was seldom used for such purpose.

Comparing with the docked bike sharing data, the dockless bike sharing data is able to record the time and location indicating the travel origin and destination in a more accurate way, providing valuable information for modelling human mobility patterns in urban systems (Yuan et al., 2012).

In order to understand how much air quality affects human mobility patterns, the air quality was measured using air quality index (AQI) and human mobility patterns were measured by travel volume and travel distance of shared bikes, the correlation between which was quantitatively computed by the Spearman Correlation Coefficient (SCC). A case study was conducted in Beijing, China to investigate the impact of air quality on human mobility patterns from spatial and temporal perspectives.

The rest of the paper is organized as follows. Section 2 introduces the study area and data used in this study. In Section 3, the main methods used are presented. Section 4 analyzes and discusses the results and some conclusions are drawn in Section 5.

### 2. STUDY AREA AND DATA

#### 2.1 Study Area

Beijing, as the first city in China where shared bicycles appeared, started the bike-sharing program in 2014. By 2017, there were 15 shared bicycle companies operating in Beijing, with the peak number of bicycles reaching 2.35 million and the number of registered users reaching 11 million (Li, 2018). As shown in Figure 1, a total of 16 administrative districts are located in Beijing, i.e., Dongcheng District, Xicheng District, Chaoyang District, Fengtai District, Shijingshan District, Haidian District, Shunyi District, Tongzhou District, Daxing District,

\* Corresponding author

Fangshan District, Mentougou District, Changping District, Pinggu District, Miyun District, Huairou District and Yanqing District. In this study, we did not include the Yanqing District as the study area because there is limited data that is not enough for analysis.

## 2.2 Data

### 2.2.1 Bike sharing data

The bike sharing data used in this study was obtained from the 2017 Mobike algorithm challenge <sup>(1)</sup>. The data ranging from May 10 to May 16, 2017 was collected, including 1.8 million records, 430,000 bicycles, and 310,000 users. It includes the fields such as order ID, user ID, bicycle ID, bicycle type, start time, start location, and destination location (see Table 1), providing useful information for the spatiotemporal analysis of human mobility patterns. The start location and destination location are originally encoded using Geohash, which can be converted to latitude and longitude in an automatic manner. More details regarding such conversion process are elaborated in Section 3.1. The kernel density of the travel volume was calculated and visualized in Figure 1, showing a tendency to spread from the central area to the rural area. The darker the region, the more the bike-sharing orders. Fengtai District, Chaoyang District, Haidian District, Dongcheng District, and Xicheng District are central areas where dense human movement occurs and have higher travel volume.

### 2.2.2 Air quality data

There are 35 monitoring stations across Beijing, which are divided into 12 urban environmental monitoring stations, 11 suburban environmental monitoring stations, 7 control and regional stations, and 5 traffic pollution monitoring points. In

this study, the spatial join function was used to map the monitoring stations to each district. There are one or two monitoring stations located in each district (i.e., the red points in Figure 1). The stations distribute in various districts, to monitor the air quality condition in each district. The air quality data was collected from the Beijing Air Quality Historical Data website <sup>(2)</sup> including the name of monitoring stations, a pair of coordinates, hourly concentrations of PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and AQI. As AQI is determined by the concentration value and sub-index of various pollutants, we selected AQI as the indicator of air quality in this study. AQI ranging from 0 to 50 represents good air quality with little or no health impact. AQI ranging from 51 to 100 represents moderate air quality with acceptable health impact for most people. AQI ranging from 101 to 150 represents unhealthy air quality for the sensitive groups such as children, the elderly, and people with respiratory diseases. AQI ranging from 151 to 200 represents unhealthy air quality for everyone, and everyone may begin to experience some adverse health problems. AQI ranging from 201 to 300 represents very unhealthy air quality, and everyone may experience more serious disease. AQI over 300 represents hazardous air quality, and everyone should avoid all outdoor activities.

## 3. METHODOLOGY

In this study, the travel volume and travel distance were calculated to indicate the human mobility patterns, and the average hourly AQI value of the monitoring stations within the district was used to represent the hourly air quality of the district. In order to explore how air quality affects human mobility patterns, a quantitative correlation analysis was conducted between AQI and travel volume as well as travel distance through computing the SCC.

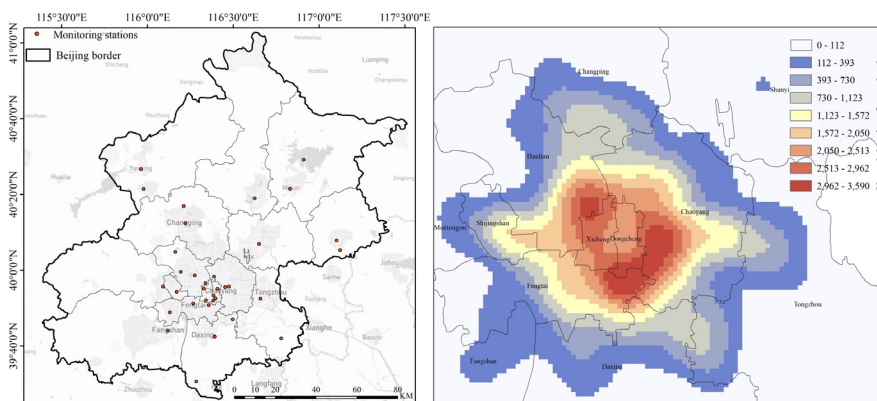


Figure 1. The study area of Beijing with the density distribution of travel volume.

Table 1. Samples of bike sharing data.

Orderid	Userid	Bikeid	Biketype	Starttime	Geohashed_Start_Loc	Geohashed_End_Loc
372218	538740	51570	1	2017/5/13 0:00	wx4eqez	wx4eqsy
193508	407216	27516	1	2017/5/13 10:11	wx4f4xd	wx4f4y3
2926825	974477	306899	1	2017/5/13 10:11	wx4fbjs	wx4dzvj
620336	595204	83778	1	2017/5/13 10:11	wx4g6yb	wx4g6wt
1536192	458250	176462	2	2017/5/13 10:11	wx4f9m2	wx4f9hf

<sup>(1)</sup> <https://www.biendata.xyz/competition/mobike/data/>

<sup>(2)</sup> <https://quotsoft.net/air/>

### 3.1 Measuring human mobility patterns

As indicated above, the human mobility patterns can be measured by travel volume and travel distance. The travel volume was computed by counting the number of orders that take place within a certain space and time. In order to investigate the actual location of bike sharing orders, the start location and destination location with geohash code as shown in Table 1 were all converted to pairs of longitude and latitude using the "geohash\_decode" function that is embedded in the Python package of Transbigdata<sup>(1)</sup>. With regard to computing the travel distance, the longitude and latitude are further converted to X and Y in meters using the "get\_distance" function embedded in the Transbigdata Python package. The travel distance can be then computed using the returned coordinates as follows,

$$Dis = \sqrt{(X_o - X_d)^2 + (Y_o - Y_d)^2} \quad (1)$$

where  $(X_o, Y_o)$  is the coordinate of the start location, and  $(X_d, Y_d)$  is the coordinate of the destination location. The air quality was measured by AQI value, which is usually used as a comprehensive indicator with consideration of various pollutants.

### 3.2 Measuring air quality

The hourly air quality of a certain district was measured by averaging the AQI values collected by the monitoring stations within in this district. Assume there are M monitoring stations within the district D, the air quality within this hour and district  $A_D$  can be computed as follows,

$$A_D = \frac{A_1 + A_2 + \dots + A_M}{M} \quad (2)$$

where  $A_1, A_2, \dots, A_M$  are the AQI values collected by the 1<sup>th</sup>, 2<sup>th</sup>, ..., M<sup>th</sup> monitoring station, respectively.

### 3.3 Computing SCC for spatiotemporal correlation analysis

The hourly travel volume and hourly travel distance were used as independent variables, and the AQI was used as dependent variables to quantitatively investigate the correlation between them based on computing the SCC. SCC is a popular indicator for correlation analysis, which is a non-parametric measure of the dependence of two or more variables. It can also usually be represented by  $\rho$ , they are identical and its value ranges from -1 to 1, where 1 means that the two variables are perfectly positively correlated, -1 means that the two variables are perfectly negatively correlated and 0 means that there is no correlation between the two variables. The SCC was calculated based on the rank order of the original variables. Rank order refers to the position of the variables after sorting them from low to high. It is calculated as follows,

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (3)$$

where  $x_i$  is the rank order of variable  $x$  (i.e., travel distance and travel volume),  $y_i$  is the rank order of variable  $y$  (i.e., AQI values),  $\bar{x}$  and  $\bar{y}$  are the mean of  $x$  and  $y$ , respectively.  $\rho$  is the correlation coefficient between variable  $x$  and variable  $y$ .

In this work, we use equation 2 to calculate the SCC and to better identify the difference between weekdays and weekends, we use equation 3 to calculate the SCC for weekdays and weekends. It is calculated as follows,

$$SCC = SCC_1 + SCC_2 + SCC_3 \dots + SCC_i/N \quad (4)$$

where  $SCC_i$  represents the SCC value of the  $i^{\text{th}}$  day and N is the number of days. The N equals to five and two with regard to weekdays and weekends, respectively.

## 4. RESULTS AND DISCUSSION

### 4.1 Spatiotemporal human mobility patterns

In this section, we analyzed human mobility patterns using the bike sharing data from spatial and temporal perspectives. With regard to the travel distance, we first made a statistic analysis in Figure 2 to illustrate the overall distribution of travel distance of all orders. It shows that the majority of the travel distance is within 5,000 meters. Additionally, the data was divided into weekdays and weekends to reflect the different spatiotemporal human mobility patterns during different time periods.

The hourly travel volume and hourly travel distance on weekdays and weekends are summarized using histograms in Figure 3. According to Figure 3 (a), it reveals that the travel volume is very low from 11pm to 5am on weekdays, which is due to people resting during these hours. The travel volume rapidly increases during 7am and 8am and 5pm and 6pm, aligning with the morning and evening peak hours. At noon, there is a local peak hour at 12pm, which is potentially related to the fact that people go out for lunch or leisure activities. The peak-hour patterns do not exist on weekends when the overall travel volume is significantly lower than weekdays, which indicates that bicycle travel largely serves workday commuting activities. With regard to the travel distance on weekdays and weekends that are shown in Figure 3 (b), there is no significant difference among other hours except there exist two peaks from 7am to 8am and from 5pm to 6pm on weekdays, which co-occurs with the higher travel volume. This is consistent with the time of going to work and off work.

Furthermore, we investigated the heterogeneity of human mobility patterns in space during morning peak hours (i.e., 7am to 9am on weekdays), evening peak hours (i.e., 5pm to 8pm on weekdays) and non-peak hours. The kernel density of travel volume and travel distance were computed and visualized in Figure 4. It shows that the travel volume density keeps high in the central districts including Chaoyang District, Haidian District, Xicheng District, Dongcheng District, and Fengtai District (see Figure 4 (a), (b) and (c)).

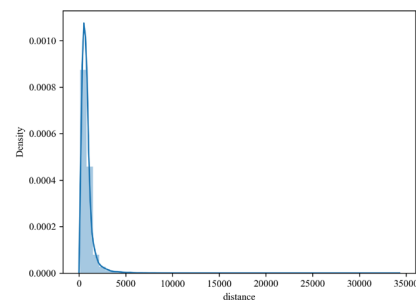


Figure 2. Statistics of travel distance.

<sup>(1)</sup> <https://transbigdata.readthedocs.io/>

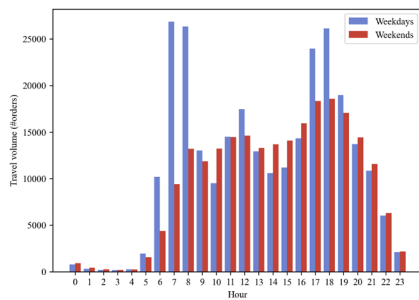
## 4.2 Air quality

Following the methodology in section 3.2, we visualized the air quality measured by the average AQI values by districts in Figure 5. It reveals that Tongzhou District, Daxing District, Fangshan District, Fengtai District and Xicheng District own higher AQI and poorer air quality during the morning peak hours on weekdays. During the evening peak hours on weekdays, the AQI is higher in Fangshan District and Mentougou District. During the non-peak hours, the AQI is the highest in Fangshan District. Overall, the air quality in Fangshan District is the poorest, regardless of the time periods.

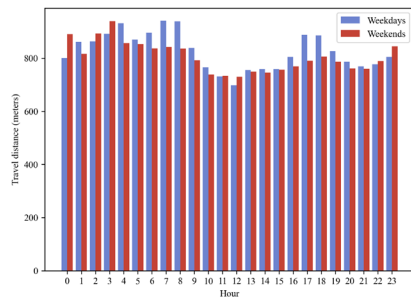
## 4.3 Correlation between spatiotemporal travel patterns and air quality

Through computing the SCC values, we quantitatively analyzed the correlation between air quality and travel volume as well as travel distance from spatial and temporal perspectives. The results are summarized in Table 2. It shows that on weekdays, the travel distance was negatively affected by air quality in areas such as Changping District (-0.20), Haidian District (-0.13), Shunyi District (-0.12), Chaoyang District (-0.10), Fangshan District (-0.10) and Tongzhou District (-0.05). Most of these areas have a high demand for shared bikes. However, the travel volume on weekdays shows randomly correlates with air quality among all administrative districts of Beijing. This is because people have to go to work regardless of the air quality on weekdays.

The travel volume has a negative relationship with AQI on weekends. The districts most affected by air quality on weekends are Changping District, Daxing District, Tongzhou District, Xicheng District, Chaoyang District, and Shijingshan District, with SCC values of -0.66, -0.58, -0.54, -0.42, -0.39, and -0.39, respectively. The negative correlation between air quality and travel volume may be due to people's reluctance to engage in outdoor activities during poor air quality. The Huairou District and Pinggu District, which are the two popular tourist attractions in Beijing that people tend to ride in spring, most align with the afore-mentioned patterns, with the SCC of -0.17, -0.16, respectively.



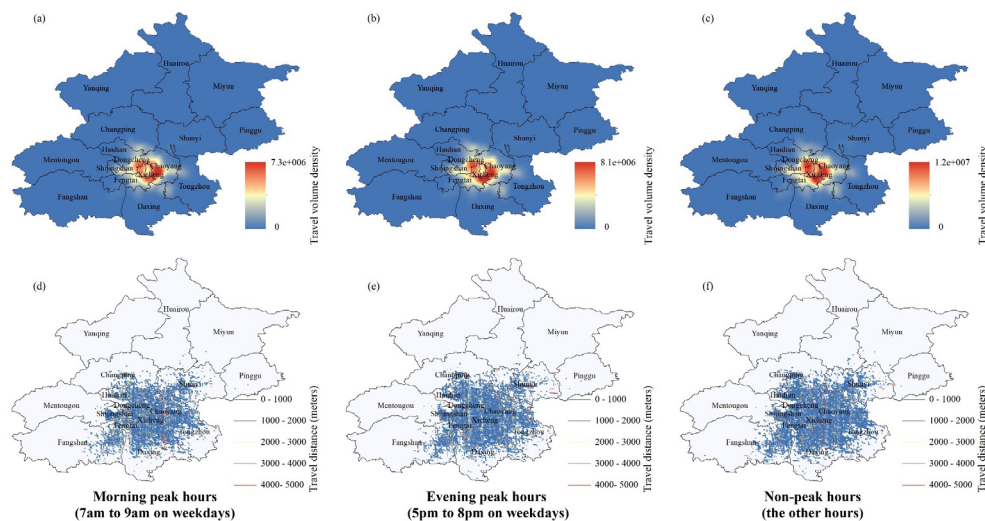
(a) Hourly travel volume



(b) Hourly travel distance

**Figure 3.** Hourly travel volume and travel distance on weekdays and weekends.

As shown in Figure 4 (d), (e) and (f), most of the travel distance of the central districts are less than two kilometers, which reflects the bike sharing system provides a solution to the "first and last mile" problem. Longer travel distance during the morning peak hours appears in Daxing District and Tongzhou District with an average distance of 3,374 meters. However, during the evening peak hours, the districts with longer travel distance drift to Chaoyang District and Haidian District, with an average distance of 3,402 meters which is slightly more than that in the morning.



**Figure 4.** Distribution of travel volume density (a-c) and travel distance (d-f) during different time intervals.

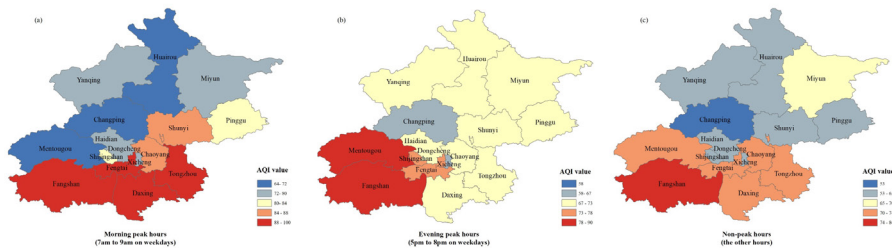


Figure 5. Distribution of AQI during different time intervals.

Table 2. SCC measuring the air quality impact on travel distance and travel volume.

District \ SCC	Weekdays		Weekends	
	Air quality impact on travel volume	Air quality impact on travel distance	Air quality impact on travel volume	Air quality impact on travel distance
Changping	0.38	-0.20	-0.66	0.09
Chaoyang	0.37	-0.10	-0.39	0.27
Daxing	0.49	0.08	-0.58	0.11
Dongcheng	0.39	0.01	-0.28	0.10
Fangshan	0.48	-0.10	-0.37	0.02
Fengtai	0.50	0.04	-0.21	-0.13
Haidian	0.51	-0.13	-0.18	0.10
Huairou	0.22	0.17	-0.16	-0.17
Mentougou	0.46	0.09	-0.19	0.01
Miyun	0.08	0.08	0.07	0.08
Pinggu	0.31	0.31	-0.15	-0.16
Shijingshan	0.39	0.06	-0.39	-0.03
Shunyi	0.40	-0.12	-0.31	0.16
Tongzhou	0.48	-0.05	-0.54	0.00
Xicheng	0.31	0.15	-0.42	-0.08

In a word, the quantitatively correlation analysis results prove that air quality indeed influences the people activity patterns. As the air quality is worse with a higher AQI value, the travel volume is less and the travel distance is shorter, where people tend to shorten their traveling by bike or keep away from going outside.

### 5. CONCLUSIONS

The travel distance and travel volume retrieved from bike sharing data were used for modelling human mobility patterns, and the AQI values were used for generating the air quality of a certain district during a specific time period. The SCC were further computed to quantitatively analyze the impact of air quality on human mobility patterns across different administrative districts of Beijing on weekdays and weekends. The results show that on weekdays, the air quality has less effect on the travel volume but worse air quality usually leads to shorter travel distance. This is because people have to follow the daily routes due to the commuting demand. On weekends, it is more flexible for people to choose their travel modes. Thus, the air quality significantly affects the travel volume and travel distance, especially in those rural districts, e.g., Changping District and Huairou District. This work helps understand human-environment coupling mechanism, promoting the harmonious development of urban systems. In the future, multiple types of crowd-sourcing data (e.g., social media and taxi data) can be fused to shape human mobility patterns and a longer series of data can be analyzed to the improve the performance.

### ACKNOWLEDGEMENT

This work was funded by the Beijing Association for Science and Technology Young Elite Scientist Sponsorship Program (BYESS2023008), the Key Laboratory of Urban Spatial Informatics, Ministry of Natural Resources of the People's Republic of China (2023ZD002), China Scholarship Council (03998521001) and Beijing Categorized Development Quota Project (03082723003).

### REFERENCES

Bagrow, J. P., & Lin, Y.-R. (2012). Mesoscopic Structure and Social Aspects of Human Mobility. *PLoS ONE*, 7, e37676.

Campbell, A. A., Cherry, C. R., Ryerson, M. S., & Yang, X. (2016). Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transportation Research Part C: Emerging Technologies*, 67, 399–414.

Cao, X., Min, J., & Huang, X. (2019). The influence of rainfall and air pollution on public bicycle usage: A case study of Xi'an. *Human Geography*, 34, 151–158.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. (2011). Exploring Millions of Footprints in Location Sharing Services. *Proceedings of the International AAAI Conference on Web and Social Media*, 5, 81–88.

Kang, C., Liu, X., Xu, X., & Qin, K. (2019). Impact of weather condition on intra-urban travel behavior: Evidence from taxi trajectory data. *Journal of Geo-Information Science*, 21, 118–127.

Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and Its Applications*, 391, 1702–1717.

Li, Y. (2018). Study on the Management Policy of Shared Bicycles in Beijing Cities. *Transport World*, 158–160.

Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106, 73–87.

Ma, L., Xiong, H., Wang, Z., & Xie, K. (2019). Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China. *Transportation Research Part D: Transport and Environment*, 68, 39–51.

Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y.-J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*, 58, 135–145.

Peng, C., Jin, X., Wong, K.-C., Shi, M., & Liò, P. (2012). Collective Human Mobility Pattern from Taxi Trips in Urban Area. *PLoS ONE*, 7, e34487.

Peng, Y., Chenxiao, Z., Mingda, Z., & Liangcun, J. (2014). Sensor Web event detection and geoprocessing over Big data. *2014 IEEE Geoscience and Remote Sensing Symposium*, 1401–1404. Quebec City, QC: IEEE.

Sagl, G., Resch, B., Hawelka, B., & Beinat, E. (2012). *From Social Sensor Data to Collective Human Behaviour Patterns: Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments*.

Wu, J., & Liao, H. (2020). Weather, travel mode choice, and impacts on subway ridership in Beijing. *Transportation Research Part A: Policy and Practice*, 135, 264–279.

Xu, Y., Liu, Y., Chang, X., & Huang, W. (2021). How does air pollution affect travel behavior? A big data field study. *Transportation Research Part D: Transport and Environment*, 99, 103007.

Yan, L., Duarte, F., Wang, D., Zheng, S., & Ratti, C. (2019). Exploring the effect of air pollution on social activity in China using geotagged social media check-in data. *Cities*, 91, 116–125.

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 186–194. Beijing China: ACM.