DTS 2,3

316

Received 20 January 2023 Revised 8 April 2023 13 April 2023 Accepted 9 May 2023

# Data-driven identification and analysis of passenger riding paths in megacity metro system

Lianghui Xie and Zhenji Zhang

School of Economics and Management, Beijing Jiaotong University, Beijing, China Robin Qiu

Department of Information Science, Pennsylvania State University, Malvern, Pennsylvania, USA, and

Daqing Gong

School of Economics and Management, Beijing Jiaotong University, Beijing, China

# Abstract

**Purpose** – The paper aims to identify and analyze passengers' riding paths for providing better operational support for digital transformation in megacity metro systems.

**Design/methodology/approach** – The authors develop a method to leverage certain passengers' deterministic riding paths to corroborate other passengers' uncertain paths. Using Automatic Fare Collection data and train schedules, a witness model is built to recover the actual riding paths for passengers whose paths are unknown otherwise. The identification and analysis of passenger riding paths between three different types of origin–destination) pairs reveal the complexity of passenger path choice.

**Findings** – The results show that passenger path choice modeling is usually characterized by complexity, experience and partial blindness. Some passengers choose paths that are not optimal due to their experience and limited access to overall metro system information. These passengers could be the subject of improved path guidance in light of riding efficiency improved through digital transformation.

**Originality/value** – This research contributes to the improvement of metro management and operations by leveraging ongoing digital transformation in megacity metro systems. Based on the riding paths and trip chains of a large number of individual passengers identified by the proposed method, metro operation management could prevent risks in areas with concentrated passenger flow in advance, optimally adjust train schedules on a daily basis and deliver real-time riding guidance station by station, which would greatly improve megacity metro systems' service safety, quality and operational efficacy over time.

Keywords Metro, Automatic Fare Collection (AFC), Path choice and identification, Witness Paper type Research paper

# 1. Introduction

With the acceleration of urbanization, the growth of traffic demand and the continuous expansion of metro network, the metro system is closely related to people's life in cities, especially in megacities. For instance, the metro system provides services for more than six



Digital Transformation and Society Vol. 2 No. 3, 2023 pp. 316-339 Emerald Publishing Limited e-ISSN: 2755-077X p-ISSN: 2755-0761 DOI 10.1108/DTS-01-2023-0006

This work was supported by the Beijing Social Science Foundation under Grant 19JDGLA002, 18JDGLA018, the MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant 19YJC630043, the National Natural Science Foundation of China under Grant J1824031 and was partially supported by the Beijing Logistics Informatics Research Base.

<sup>©</sup> Lianghui Xie, Zhenji Zhang, Robin Qiu and Daqing Gong. Published in *Digital Transformation and Society*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http:// creativecommons.org/licences/by/4.0/legalcode

million passengers on weekdays in Beijing, and the sharing rate in public transport reaches 57.4% (in 2021), making a significant contribution to convenient, efficient, green and safe travel for residents. The travel trajectories of millions of passengers in the metro system are gathered into passenger flows, and the agglomeration, dissipation and congestion of these passenger flows have become one of the most concerned contents of operation managers.

Understanding and predicting the passenger flow characteristics of a metro system in real time will help operate the system safely, cost-effectively and sustainably, and provide better and more satisfying city transportation services in the long term. The traditional method to capture passenger flow distribution characteristics was to conduct field surveys at metro stations and ask people about their destinations and route choices (Rashedi, Mahmoud, Hasnine, & Habib, 2017; Bajaj & Singh, 2021). Then, a route choice model was built, which had parameters related to travel time, cost, walking distance, the number of transfers and crowding situation (Liu, Sun, Bai, & Xu, 2009; Xu, Luo, & Gao, 2009; Raveau, Muñoz, & Grange, 2011; Jin, Yao, Zhang, & Liu, 2017). This approach is time-consuming and labor-intensive in terms of conducting surveys and processing data (Zhao *et al.*, 2017). Due to the limited sample size, its results are often error-prone and subjective.

Safety, efficiency and service are the eternal theme of metro system. After more than a century of development, the metro system has become the most safe, reliable and efficient mode of transportation for residents of megacities. It has not only devoted to the tireless innovation and exploration of metro operation managers but also has the distinctive characteristics of technological progress and concept transformation of the times. Entering the information age, digital technology is now widely recognized as an important opportunity to pioneer delicacy management and develop innovative services. Companies must constantly innovate to keep up with the current possibilities and trends, a process commonly referred to as digital transformation (Van Veldhoven & Vanthienen, 2023). Metro system has naturally become an important field of information technology and digital technology application.

With the rapid development of sensor-based networks, Automatic Fare Collection (AFC) systems have been widely adopted by metro systems. Meanwhile, mining the travel behavior of passengers from AFC data has become a trend in the field of passenger flow modeling research (Sun, Jin, Lee, Axhausen, & Erath, 2014; Zhao, Qu, Zhang, Xu, & Liu, 2017; Barry, Newhouser, Rahbee, & Sayeda, 2002; Tang, Zhao, Cabrera, Ma, & Tsui, 2019; Sun, Lu, Jin, Lee, & Axhausen, 2015; Wang, Zhang, Zhao, Liu, & Zhang, 2020; Xie, Zhang, & Gong, 2022). AFC data provide a possibility for fine understanding of temporal and spatial distribution of passenger flow. However, AFC data only record entry and exit information and do not contain any trip connection information. The trajectory of passenger in metro system is a typical "black box" problem. How to excavate passenger riding path and path selection characteristics hidden behind AFC data by data-driven method is worth further study.

The purpose of this study is to develop a new heuristic method to identify passengers' paths and analyze their riding characteristics. The remainder of the paper is organized as follows. The related works in the literature are presented in Section 2. Section 3 proposes our new research model. Section 4 provides the results of this study, which are discussed from the managerial perspective in Section 5. Finally, Section 6 concludes the paper.

#### 2. Research background

Researchers have solved the task of passenger flow distribution using a variety of different methods. According to the objective of such a task, we usually know the passenger flow distribution of a metro network roughly but do not know the riding trajectory of each passenger and his/her riding characteristics. With the application of AFC systems and the

Megacity metro system

development of big data technology, scholars have put together a lot of effort in addressing the identification problem of individuals' riding paths over the last decade or so.

Based on the assumption that a passenger's access time should be in the same percentile of the cumulative access time distribution as the egress time is in the distribution, Paul (2010) explored a method of assigning passenger trips to feasible itineraries. When there were multiple feasible itineraries, the most probable itinerary was selected according to some proposed rules. Kusakabe, Iryo, and Asakura (2010) proposed an algorithm to estimate the train that a passenger boarded using smart card data. They assumed that a passenger would minimize the total waiting time at the departure station and lost time at the arrival station and choose the route with the least number of transfers.

Sun and Xu (2012) used AFC data, train schedules and supplementary manual surveys to infer passengers' entry, exit and transfer times. Using the collected information, they developed travel time distributions and inferred route choice proportions based on the actual travel time of passengers and the time distribution of each path in their studied network. Zhou and Xu (2012) estimated an individual passenger route choice with the maximum likelihood method using AFC data and train operation information. They assumed that the walking speed of a passenger stays the same as others and that the transfer delay for an individual passenger due to crowding occurs with the same probability as others. Based on this assumption, the passenger dwell time at his/her origin station and transfer stations was derived, and the matching degree for each path according to the maximum likelihood boarding plan was calculated.

Zhu, Hu, and Huang (2014) proposed a method to calibrate an urban metro assignment model using AFC data. The method used a framework based on a genetic algorithm with nonparametric statistical technology, first generating a candidate set using statistics-based criteria and then applying a genetic algorithm to find an optimal solution. Hong, Min, Park, Kim, and Oh (2016) derived trip chain information from reference passengers whose trips are known. To detect an unknown path of a passenger, the proposed method checks if it optimally forms a sequence of boarding, transfer train(s) and alighting trains for each alternative connection.

Hörcher, Graham, and Anderson (2017) proposed a passenger-to-train assignment method that recovers the crowding density in the entire metro network. The assignment is based on the likelihood of access, transfer and egress times associated with feasible itineraries in the Automatic Vehicle Location (AVL) dataset. Xu, Xie, Li, and Qin (2018) adopted a method combining Bayesian inference and Metropolis-Hasting sampling to learn the route choice behavior of passengers from AFC, train schedules, and train loading data and calibrated the parameters of a logit model.

Zhang, Yao, Zhang, and Zheng (2018) proposed a set of approaches for estimating passenger paths by combining self-reported revealed preference (RP) and smart card data. A nested model was developed with a balance parameter by accommodating different scales of the two data sets. Combining the advantages of RP and AFC data on a large scale, the model performed better than a single data source through. Li, Luo, Cai, and Zhang (2018), and Wu *et al.* (2019) used the clustering method to group the travel times of passengers between OD pairs into different clusters. A method considering both uncertain walk times and transfer times was proposed to estimate the theoretical travel times of all possible paths.

Zhu, Koutsopoulos, and Wilson (2021) developed the Passenger Itinerary Inference Model (PIIM). After calculating the probabilities of left behind and route choice respectively in multiple paths of origin–destination (OD) pairs, the probability of passengers matching each trip chain was calculated by Bayesian inference. Su, Si, Zhao, and Li (2022) selected the AFC data between OD of special types to estimate the distribution of platform passengers' walking time and waiting time first. They then established the real-time path travel time distribution

DTS 2,3

with the distribution of walking time, waiting time and in-vehicle time as parameters. Finally, the membership function is introduced to evaluate the relationship between the passenger travel time and the real-time travel time distribution of each candidate path, and the path with the largest membership degree is taken as the passenger riding path.

Regardless of the adopted methods over the years, the results from the existing studies reviewed above mainly show the likelihood of multiple paths. Little study has been done in terms of enabling a real-time applicable method to identify passengers' paths and understand their riding behaviors. In addition, the above method based on some assumptions, such as passengers will not be left behind, there is no difference in the behavior of passengers in a day, and the tap-out time of passengers who get off from two consecutive trains does not overlap. They are inconsistent with reality, affecting the accuracy of the results and limiting its applicability. Therefore, passenger riding path identification methods that can be applied to large-scale metro networks still need to be developed.

## 3. Investigation methodology

# 3.1 Fundamental principles

A typical urban metro trip consists of sub-paths or segments. As shown in Figure 1, sub-paths of a passenger's trip are clearly described by the following activities: ① tap-in, ② walking to the platform and waiting for his/her train, ③ riding, ④ transferring from the train and waiting for his/her next train, ⑤ riding, ⑥ walking to the gates and ⑦ tap-out.

Finding the riding time in each of the sub-paths described above is essential for uncovering his/her riding path. The timestamps at ① and ⑦ can be directly obtained from AFC data. The durations of ③ and ⑤ can be obtained from train schedules if his/her transfers are uncovered. The duration of ⑥ cannot be directly determined because the walking speed of a passenger is unknown. The durations of ② and ④ consist of both unknown walk and waiting time. Therefore, it is challenging to determine the actual riding path of a trip if the time of each sub-path along the trip is not identified. To uncover the riding paths of passengers, the following questions must be answered.

*Q1.* Is it possible to determine the paths taken by certain people?

Even in a highly complex metro network, there are some passengers whose riding paths can be determined. For example, on March 26, 2018, passenger A tapped in at CGZX at 17:04:00 and exited from HDWLJ at 17:16:57. Because of the direct access to CGZX and HDWLJ via Line 6, as shown in Figure 2, there is a good reason to believe that the passenger took Line 6 to HDWLJ. The passenger's riding time was 12 minutes and 57 seconds; no other paths could have been completed within this time frame, which verifies that the uncovered passenger's riding path was correct.

Q2. Is it possible to determine the train that certain people take?

Once the path of a passenger is identified, train schedules make it possible to determine the train that the passenger took. We continue to use passenger A discussed above as an example. There were three trains near passenger A's tap-in time, as shown in Table 1.



Figure 1. Typical metro trip composition

Megacity metro system



	Station	No. 1209	No. 1210	No. 1211
	CGZX (Departure) BSON (Arrival)	17:02:38	17:05:22	17:08:06
	BSQN (Departure)	17:05:28	17:07:22 17:08:12 17:00-52	17:10:56
	HYQ (Arrival) HYQ (Departure)	17:07:08 17:07:38	17:09:52	17:12:36
Table 1.	CSS (Arrival) CSS (Departure)	17:09:32 17:10:17	17:12:16 17:13:01	17:15:00 17:15:45
Frain schedule from CGZX to HDWLJ	HDWLJ (Arrival) Source(s): Table by the authors	17:12:31	17:15:15	17:17:59

Train No. 1209 departed from CGZX at 17:02:38, which was earlier than passenger A's tap-in time; thus, passenger A did not take this train. Train No. 1211 arrived at HDWLJ at 17:17:59, which is later than passenger A's tap-out time; thus, passenger A did not take this train either. Therefore, passenger A must have been on the only train that meets the requirement, i.e. No. 1210.

*Q3.* Is it possible to determine the time spent in each sub-path in the entire trip of a passenger with an identified path with identified trains?

After a path and train numbers (or lines) are identified, the time of sub-paths ①, ③, ⑤ and ⑦ can be respectively determined, and ②, ④ and ⑥ are exactly in the middle of the known segments, so their time can be calculated. For example, passenger A tapped in at CGZX at 17:04:00, then walked to the platform and waited for the train for 1 minute 22 seconds, took Line 6 at 17:05:22, got off at 17:15:15 after riding for 9 minutes 57 seconds without a transfer, then walked to the gate for 1 minute 42 seconds, and tapped out at HDWLJ at 17:16:57. In this way, we uncovered the complete trip information of passenger A, as shown in Figure 3.

Q4. Can we use passengers with known trips to make proofs for other passengers?

The full riding history of passenger A is not only meaningful for analyzing the riding behavior of passenger A but also reveals important information that might help us analyze the riding paths of other passengers. The revealed information includes the following: (1) passengers who tapped in at CGZX before 17:04:00 would take No. 1210 train that departed at 17:05:22; (2) after tap-in at CGZX, passengers would take about 1 minute and 22 seconds to walk to the platform; (3) passengers would take 1 minute and 42 seconds to walk out and then tap-out at HDWLJ after getting off the train. For passenger B, who tapped in at CGZX before 17:04:00, she could have also taken No. 1210 train. Although passenger A and passenger B did not know each other, heuristically, passenger A could be considered as a "witness" who could testify that passenger B took the same train with him/her.

In this study, passenger A is heuristically defined as a "witness." Many "witnesses" like passenger A with known paths can be identified. Some of these witnesses could witness that target passenger B entered a given station and then took a given train. Some could witness that passenger B took a given train and then exited a given station. Others witness that passenger B took a given train and then transferred to another given train.

#### 3.2 Assumptions

In this study, we have the following assumptions that form the basis of the subsequent algorithms.

Assumption 1. Once passengers get off a train, they will leave the train station right away.

There are many factors affecting the egress time of passengers. Each passenger has a different walk speed, and the different locations of the carriages lead to different walk distances, making the same wave of passengers exit the station at different times; however, passengers will generally leave the station as soon as possible after getting off. According to previous studies, the distribution of the egress time of passengers in the same wave obeys the extreme value distribution (Hong, Min, Park, Kim, & Oh, 2016). We can easily know from which train the exiting passengers got off, as shown in Figure 4.

Assumption 2. When a path is the shortest path and the only direct path between a certain origin and destination, passengers will definitely choose this path.

In a metro network, there may be an infinite number of reachable paths between an OD pair but only a small number of efficient paths, i.e. people will generally choose only efficient paths with a lower impedance utility rather than invalid paths with features such as detours, long riding times and many transfers. When there is only one path in an OD pair in a set of efficient paths, passengers will uniquely choose that path. In other words, when a path is the shortest path and the only direct path between a certain origin and destination, passengers will definitely choose this path, which has an absolute advantage in terms of riding time and the number of transfers.

*Assumption 3.* When there is no direct path between an origin and destination and a path is the shortest path and the only path with one transfer, passengers will definitely choose this path.



Figure 3. Schematic diagram of passenger A's trip recovery

Megacity metro system

DTS 2,3

322



**Figure 4.** Tap-out flows at a metro station When there is no direct path between an OD pair, adding one transfer to Assumption 2, we can infer that if a path is the shortest path and the only path with one transfer, passengers will definitely choose this path. This assumption offers absolute advantages in terms of riding time and the number of transfers.

#### 3.3 Definitions

We have the following definitions.

#### Definition 1. Trip chain—a combination of basic information about a passenger's trip.

The trip chain includes the time and stations of passenger's tap-in and tap-out, the train numbers of each train he/she take during the journey, as well as departure and arrival time of each train.

# Definition 2. Class I witness $W_1$ —nontransfer passengers whose riding paths and trip chains are determined.

According to Assumption 1, a passenger exits his/her destination station without stopping after getting off the train. Therefore, according to the time of exit, the nontransfer passenger's train number can be determined, which satisfies the requirement that the "trip chain is determined." According to Assumption 2, when a path is the shortest path and the only direct path, passengers will definitely choose this path. Therefore, the nontransfer passengers who take these paths can satisfy the requirement that the "riding path is determined" and are Class I witnesses.

# Definition 3. Class II Witness $W_2$ —one-transfer passengers whose riding paths are determined.

According to Assumption 3, when there is no direct path between an origin and destination and a path is the shortest path and the only path with one transfer, passengers will surely choose this path. Therefore, these one-transfer passengers can satisfy the requirement that "riding paths are determined" and are Class II witnesses.

According to Assumption 1, the train numbers of Class II witnesses can be determined when they exit the station; thus, the train numbers before the transfer can be inferred. However, in a rush-hour congestion condition, passengers may be left behind and need to wait for multiple trains, in which case we cannot determine whether the passengers are left behind before or after the transfer. Therefore, when the train numbers of Class II witnesses before a transfer cannot be determined, all their possible train chains are marked, which is also very useful for subsequent corroboration of the target passenger's transfer behavior.

# Definition 4. Minimum egress time $T_d^{egr}$ —the minimum time to walk to tap-out after alighting at station d.

The egress time of a large number of Class I witnesses during off-peak hours at station d can be assessed. The minimum time is  $T_d^{egr}$ , which means that the passengers exiting the station have the shortest walk distance and the fastest walk speed. Taking the moment of train arrival plus the minimum egress time as the splitting point, the witnesses who exit the station at any moment can be distinguished, and the train number they took can be delineated.

# Definition 5. Minimum access time $T_o^{acc}$ —the minimum time to board a train after tap-in at station o.

The access time of a large number of Class I witnesses at station o can be assessed too. The minimum time is  $T_o^{acc}$ , which corresponds to the time of passengers entering the station with the shortest walk distance, the fastest walk speed and no waiting time to board the train.

323

Megacity

metro system

Definition 6. Minimum transfer time  $T_s^{tra}$ —the minimum time to transfer at station s.

The transfer time of a large number of Class II witnesses at station *s* can be assessed counted. The minimum time is  $T_s^{tra}$ , which corresponds to the time that passengers transfer with the shortest walk distance, the fastest walk speed and without waiting for the train to transfer.

Definition 7. Minimum riding time  $T_i^{\min}$ —the minimum time to ride on a certain path j.

 $T_j^{\min}$  is the sum of the minimum access time, minimum transfer time, minimum egress time and on-board time on path *j*, which corresponds to the time that passengers enter an original station, ride on trains and exit a destination station with the shortest distance, the fastest walk speed and no waiting time to board the transfer trains.

# Definition 8. Search scope of witness S—the search scope for witnesses that target passengers may encounter at the time when those target passengers enter, exit and transfer in their trips.

The search scope of Class I witnesses when the target passenger k enters is  $[t_k^{in} - T_d/2, t_k^{in} + T_d/2]$ , where  $t_k^{in}$  is the tap-in time of passenger k, and  $T_d$  is the departure interval of trains at the time. The search scope of Class I witnesses when the target passenger k exits is  $[t_k^{out} - T_d/2, t_k^{out} + T_d/2]$ , where  $t_k^{out}$  is the tap-out time of passenger k.

The search scope of Class II witnesses when the target passenger k transfers: Once Class I witnesses at the exit are identified, their train numbers are also determined. Regardless of the origin and destination of a Class II witness, a Class II witness who transfers to the train number of a Class I witness within the possible riding path of the target passenger is a Class II witness along the transfer path. When the number of transfers is greater than 1, the train number determined by the Class II witnesses of the last transfer is used as the basis for looking for witnesses of the previous transfer. An example is shown in Figure 5. When the train number of Class I witnesses at the exit is determined to be N1, then a Class II witness at the second transfer should be transferred to train N1. Suppose the train chains of the Class II witness at the second transfer are (M1, N1) and (M2, N1); then, the Class II witness at the first transfer should be transferred to train M1 or M2, such as the one whose train chain is (L1, M1) or (L2, M2). However, the Class II witness whose train chain is (L3, M3) or (M2, N2) is not the witness of the target passenger.

#### 3.4 Algorithm

In the first step, a set of efficient paths R is generated for a given OD pair; here, the paths recommended by Baidu Maps are directly used as efficient paths, which are alternative paths that passengers usually choose on a daily basis.



Figure 5. Schematic diagram of the search scope of Class II witnesses

DTS 2.3 The second step is to judge whether the riding time of a target passenger is smaller than the minimum riding time of each path, where if the riding time of the target passenger is smaller than the minimum riding time of a path, then the path is invalid.

The third step is to determine whether the path can form valid trip chain, i.e. there are witnesses at the time of entry, transfer, and exit and tandem train numbers of these witnesses riding. For example, there is a valid trip chain whose train chain is (L2, M2, N1), as shown in Figure 6.

The fourth step is to determine the number of valid trip chains and paths. If the number of valid path is equal to 1, then the path is the only possible riding path for the target passenger; If the number of valid paths is greater than 1, then the voting mechanism is triggered, and the trip chain with the largest number of votes are taken as the most-likely trip chain for the target passenger, and its path is the most-likely riding path.

The voting mechanism is as follows.

- (1) Both Class I witnesses and Class II witnesses have voting rights.
- The closer the tap-in or tap-out time between a Class I witness and the target (2)passenger, the greater the weight of the witness' vote is. Let the time interval between Class I witness x and the target passenger be  $T_x$ ; then, his/her voting weight is  $Poll_x = 1/(T_x + 1).$
- (3) All Class II witnesses vote bear the weight of 1.
- (4) Let the total number of votes of inbound witnesses in the *j* trip chain of path *i* be  $W_{ii}^{in}$ the total number of votes of outbound witnesses be  $W_{ij}^{out}$ , and the total number of votes of transfer witnesses be  $W_{ij}^{tra}$ . Additionally, let the total number of votes of inbound witnesses on path *i* be  $W_i^{in}$ , the total number of votes of outbound witnesses be  $W_i^{out}$ , and the total number of votes of transfer witnesses be  $W_i^{tra}$ . Then, the voting result  $V_{ij}$  for the *j* trip chain of path *i* is:

$$V_{ij} = \frac{W_{ij}^{in} W_{ij}^{tra} W_{ij}^{out}}{W_{i}^{in} W_{i}^{tra} W_{out}^{out}}$$
(1)

The voting algorithm is shown using a flowchart in Figure 7.



Megacity metro system

325

Figure 6.



# 4. Results and findings

4.1 Data description

The Beijing Metro system is used as an application case study in this research, which has a complex network and has a strong representation. By the end of 2022, there are 27 lines in operation in Beijing, with an operating mileage of 797.3 km. There are 472 stations, including 78 transfer stations, as shown in Figure 8.





Source(s): Figure by the authors

# (1) AFC data

The data recorded by Beijing Metro AFC system is the full sample data, which contain the travel information of all urban rail transit passengers (including light rail, maglev and other lines). Taking the data from March 26, 2018 to March 29, 2018 as an example, the data volume is 26,065,883, with about 6.5 million daily trips.

The Beijing Metro AFC system records the transaction information of passengers on a single trip, including card ID, tap-in time, tap-out time, boarding station ID, alighting station ID, card type etc., as shown in Table 2.

Due to the diversity of AFC related equipment manufacturers and the complexity of AFC real-time data transmission, AFC data sometimes have recording exceptions, such problems mainly include recording errors, record missing, record duplication, etc. In addition, there are some passengers begging, stealing, advertising and delivering express in metro system, and their data should be eliminated. Therefore, prior to the analysis, data cleaning should be carried out on the AFC data, and the following data should be removed according to experience:

- There are errors, omissions or duplicates in the records.
- Tap-in and tap-out stations are the same.
- The travel time between OD pair is more than twice the minimum travel time or more than 30 minutes longer than the minimum travel time (Si, Zhong, Liu, Gao, & Wu, 2013; Hong *et al.*, 2016).

After data cleaning, 24,034,806 pieces of the above data remain, and the exclusion ratio is about 7.8%.

(2) Train schedule data

Train schedule data includes the arrival time and departure time of each train at each station. Table 3 lists the main information. When the metro system fails, the actual train arrival and departure time will be different from the train schedule, AVL data can be used instead of the train schedule data.

#### 4.2 Method validation

In order to verify the accuracy of the method, we recruited volunteers to take the Beijing Metro according to the set path, and then extracted the volunteers' travel information from

Name	Description	
CARD_SERIAL_NUMBER ENTRY_TIME TRIP_ORIGIN_LOCATION TXN_DATE_TIME DEVICE_LOCATION PRODUCT_ISSUER_ID PRODUCT_TYPE PAYMENT_VALUE RECONCILIATION_DATE SETTLEMENT_DATE DEVICE_ID SOURCE_PARTICIPANT_ID <b>Source(s):</b> Table by the authors	Card ID (encrypted) Tap-in time (year/month/day/hour/minute/second) Boarding station ID Tap-out time (year/month/day/hour/minute/second) Alighting station ID Card type, including E-card and one-way card Card subtype, including ordinary card, senior card, student card etc. Deduction amount Time for reconciliation of accounts Time for clearing currency receipts and payments Automatic fare gate ID Operating company ID	<b>Table 2.</b> Main information of AFC data in Beijing Metro system

Megacity metro system

DTS 2,3	NO.102013	06:09:48 06:09-18	06:06:58	06:06:28	06:04:05	02:00:00	06:00:35	
328	NO.012012	06:06:48 06:06:18	06:03:58	06:03:28	06:01:05	00:00:20	05:57:35	
	NO.082011	06:01:58 06:01:28	05:59:08	05:58:38	05:56:15	06.65.50	05:52:45	
	NO.072010	05:57:58 05:57:28	05:55:08	05:54:38	05:52:15	00:TC:CO	05:49:30	
	NO.062009	05:54:28 05:53:58	05:51:38	05:51:08	05:48:45 05:49:00	00:46:00	05:45:15	
	NO.052008	05:50:04 05:49:40	05:47:10	05:46:40	05:44:10	05:43:40	05:40:55	
	NO.042006	05.41.06 05.40.36	05:38:06	05:37:36	05:35:06 05:34:06	00:40:00 20:00:20	05:31:51	01S
	lber	Departure Arrival	Departure	Arrival	Departure	Development	Departure Arrival	: Table by the auth
Table 3.         Train schedule of       Beijing Metro (sample)	Train num	CHS	HS		DWL		GM	Source(s)

the AFC system. The volunteers' riding paths were identified according to the proposed method and the existing methods, including PIIM (Zhu, Koutsopoulos, & Wilson, 2021) and RP (Hong *et al.*, 2016), and then compared and verified with the actual riding paths. For specific experimental design, see Xie, Zhang, Qiu, Gong, and Ma (2023).

The results show that the method can accurately identify 141 out of 143 normal trips, with an accuracy of 98.6%. It is better than PIIM and RP methods, whose accuracy is 85.3% and 91.6%, respectively, in this experiment. This method has significant advantages in riding path identification, especially for paths with multiple transfers, the accuracy will not decrease due to the uncertainty of waiting time for transfer. On the contrary, due to the superposition of multiple uncertainties, it is not easy to form valid trip chains on other paths, so the results are more inclined to the most actual riding path.

#### 4.3 Path choice results

The following analysis uses Beijing Metro AFC data from March 26 to March 29, 2018. We selected three OD pairs: from XS (Xi Si Station) to BJXZ (Bei Jing Xi Zhan Station), from CSS (Ci Shou Si Station) to XD (Xi Dan Station) and from CGZX (Che Gong Zhuang Xi Station) to CWM (Chong Wen Men Station), as shown in Figure 9.

From Figure 10, we can see that XS and CGZX have typical business-type characteristics, with more passengers exiting the station in the morning peak and entering the station in the evening peak. CSS has typical residential-type characteristics, with more passengers entering the station in the morning peak and exiting the station in the evening peak. CWM and XD are general-type stations have typical city mass flow morning and evening peaks. BJXZ is a hub-type station, with a balanced flow of passengers in and out of the station throughout the day, and no obvious morning and evening peaks.

Then we try to analyze the performance of path choice for the same OD pair and the differences between different types of OD pairs.

### (3) CGZX—CWM

There are two efficient paths between CGZX with CWM:  $\textcircled{}{} \leftrightarrow \textcircled{} and \textcircled{}{} \leftrightarrow \textcircled{}$ , both involve one transfer and have a total of 8 stops.

There were 379 trips from CGZX to CWM during the designated period in this analysis. Path identification results show that 56.2% of passengers using path  $\textcircled{o} \rightarrow \textcircled{O}$ , with an



**Source(s):** Figure by the authors

Figure 9. Schematic view of studied metro lines

Megacity metro system



average riding time of 27 minutes and 21 seconds, and 43.8% of passengers using path ( $\textcircled{O} \rightarrow \textcircled{O}$ ), with an average riding time of 30 minutes and 20 seconds, as shown in Table 4. For 435 trips from CWM to CGZX, 75.9% of them chose path ( $\textcircled{O} \rightarrow \textcircled{O}$ ), with an average riding time of 26 minutes and 42 seconds, and others chose path ( $\textcircled{O} \rightarrow \textcircled{O}$ ), with an average riding time of 30 minutes and 24 seconds.

(4) CCS—XD

There are two efficient paths between CCS with XD: ( $\textcircled{O} \leftrightarrow \textcircled{O}$  and ( $\textcircled{O} \leftrightarrow \textcircled{O}$ ). The former has 8 stops, the latter has 7 stops, and both of them involve one transfer.

There were 517 trips from CCS to XD during the designated period in this analysis. Path identification results show that 46.8% of passengers using path  $\textcircled{O} \rightarrow \textcircled{O}$ , with an average riding time of 27 minutes and 35 seconds, and 53.2% of passengers using path  $\textcircled{O} \rightarrow \textcircled{O}$ , with an average riding time of 26 minutes and 15 seconds, as shown in Table 5. For 447 trips from XD to CCS, 54.8% of them chose path  $\textcircled{O} \rightarrow \textcircled{O}$ , with an average riding time of 28 minutes and 18 seconds, and others chose path  $\textcircled{O} \rightarrow \textcircled{O}$ , with an average riding time of 27 minutes and 40 seconds.

(5) XS—BJXZ

There are two efficient paths between XS with BJXZ:  $(4) \leftrightarrow (7)$  and  $(4) \leftrightarrow (9)$ . The former has 8 stops, the latter has 9 stops, and both of them involve one transfer.

	OD	Paths	Number of stops	Number of transfers	Average riding time (ATT, min:sec)	Number of trips	Proportion (%)
<b>Table 4.</b> Path choice resultsbetween CGZXand CWM	CGZX to CWM CWM to CGZX		8 8 8 8	1 1 1 1	27:21 30:20 26:42 30:24	213 166 330 105	56.2 43.8 75.9 24.1

There were 688 trips from XS to BJXZ during the designated period in this analysis. Path identification results show that 48.8% of passengers using path  $\textcircled{4} \rightarrow \textcircled{0}$ , with an average riding time of 29 minutes and 2 seconds, and 51.2% of passengers using path  $\textcircled{4} \rightarrow \textcircled{0}$  with an average riding time of 31 minutes and 22 seconds, as shown in Table 6. For 499 trips from BJXZ to XS, 67.1% of them chose path  $\textcircled{0} \rightarrow \textcircled{4}$ , with an average riding time of 31 minutes and 32 seconds, and others chose path  $\textcircled{0} \rightarrow \textcircled{4}$ , with an average riding time of 31 minutes and 39 seconds.

#### 4.4 Detailed analysis

The above results of passenger path choice behavior reveal that the passenger path choice behavior varies greatly among OD pairs. For further analysis, we define the frequent rider (FT), who takes the metro every day during the designated period in this analysis and divides the period into morning peak, evening peak and off-peak hours.

As can be seen in Figure 11, there are differences in the traffic distribution characteristics between the OD pairs based on different station attributes, with most passengers between CGZX and CWM concentrated in the morning and evening peaks, and a clear tidal pattern between CSS and XD, with most of them riding from CSS to XD in the morning peak and returning in the evening peak. Passengers from the XS and BJXZ OD pairs are more evenly distributed throughout the day, with no significant peaks. Accordingly, there are almost no frequent riders between XS and BJXZ, while frequent riders of other OD pairs mainly appear in the morning and evening peak hours.

After the comparative analysis, we have the following findings:

- (1) The average riding time of frequent riders is shorter than that of non-frequent riders on the same path during the same period, which may be due to the fact that frequent riders are more familiar with the riding conditions.
- (2) Frequent riders are better at choosing paths than other passengers, and a higher proportion of them choose the path with the shortest average riding time. For

OD	Paths	Number of stops	Number of transfers	Average riding time (ATT, min:sec)	Number of trips	Proportion (%)	
CSS to XD	$(6) \rightarrow (4)$ $(0) \rightarrow (1)$	8 7	1	27:35 26:15	242 275	46.8 53.2	
XD to CSS	(d)→(b) (1)→(t)	8 7	1	28:18 27:40	245 202	54.8 45.2	Path choice results
Source	(s): Table b	by the authors					and CWM

OD	Paths	Number of stops	Number of transfers	Average riding time (ATT, min:sec)	Number of trips	Proportion (%)	
XS to	(4)→⑦	8	1	29:02	336	48.8	
BJXZ	(4)→(9)	9	1	31:22	352	51.2	
BJXZ to	(7)→④	8	1	28:32	335	67.1	Table 6
XŠ Source(s)	(9)→④ Table by	9 y the authors	1	31:39	164	32.9	Path choice result between XS and BJX2

Megacity

metro system

DTS 2,3



Figure 11. Scatter diagram of riding time



DTS 2,3

334



Figure 11.

example, in Figure 11(a), 83.3% of frequent riders chose the shortest path, higher than other passengers (73.7%). In Figure 11(b), 63.0% of frequent riders chose the shortest path, higher than other passengers (54.6%).

- (3) Passengers have very complicated reasons for making path choices. For example, more than half of the passengers who traveled from XS to BJXZ chose ④→⑨, although this path had more stops and a longer average riding time compared to the other path. One possible reason for this is the transfer station GJTSG at ④→⑨, which is the departure station of Line 9 and passengers are more likely to have seats.
- (4) The path chosen by round-trip passengers of the same OD pair also varies with a variety of reasons. For example, during peak hours, 24.1% of passengers from CWM to CGZX chose path (⑤→⑥), while 43.8% of the passengers from CGZX to CWM chose path (⑥→⑤). We know that whether from CWM to CGZX or from CGZX to CWM, the number of stops does not change, and it would not be crowded during off-peak hours. It is obviously difficult to accurately estimate the choice proportion with the traditional route choice model, because its parameters do not change, and the estimated choice proportion does not change either.
- (5) Due to the difference in individual walking time and the randomness of waiting time (especially the transfer waiting time), the riding time of different paths between the same OD pair overlaps to a large extent. For example, for passengers riding from CSS to XD, although the average riding time for path (()→④) is 2 minutes and 39 seconds longer than that for path (()→①) during off-peak hours, there are still many passengers riding by path (()→④) who spend less time than path (()→①). This is because sometimes a passenger who chooses path (()→①) may have to wait longer for both Line 10 and Line 1.

Table 7 shows detailed data on the riding path choices of frequent and non-frequent riders for each period.

#### 5. Discussions

In previous studies, the focus has been more on the possibility of passengers choosing different paths to derive the overall passenger flow distribution, which plays an important role in metro network planning, schedule optimization and passenger flow diversion when needed. With the advancement of data collection, storage and analysis technologies, tools and methods have made uncovering individual riding paths possible in metro systems. For example, many scholars have made many attempts to identify individual riding paths, e.g. no left behind (Zhou, Shi, and Xu, 2015; Li, Luo, Cai, & Zhang, 2018) and stable left behind (Zhou and Xu, 2012; Zhao, Qu, Zhang, Xu, & Liu, 2017, Zhao *et al.*, 2017; Zhu *et al.*, 2021). However, the results from those attempts are not practically applicable for implementation in practice. According to our statistics, less than 1.78% of passengers in this study may have the behavior of waiting for friends, going to the bathroom etc. Hence, assumptions in this study were practically rational. More convincingly, we proposed the witness voting method. Note that if some witnesses did not comply with our assumptions during their trips, their votes would be overwhelmed by the other majority of witnesses.

Different from other methods whose accuracy decreases as the metro network becomes complex, this method has more obvious advantages in complex network. Since it requires witnesses to be found at all phases of a journey, the more transfers there are, the less likely it is to form a valid trip chain on the wrong path.

Since witnesses play a key role in the method, the method may fail when there is no witness at a certain phase of the passenger journey. This situation mainly occurs in suburban

Megacity metro system

DTS 2,3	ATT	0:24:47 0:27:19 0:27:16 0:27:16	0:26:22 0:27:37 0:28:23 0:28:23	0:26:57 0:28:39 0:27:47 0:27:30	0:26:18 0:28:13 0:25:01 0:25:01	0.26:32 0.29:05 0.28:58 0.31:24	0:28:32 0:31:39	
336	Total Proportion (%)	18.4 57.5 3.7 20.5	12.1 44.1 86 7	43.4 6.5 8 7	15.5 15.5 37.7	47.8 0.7 50.4	67.1 - 32.9	
	ATT	0:27:28 0:29:22 - 0:33:43	0:26:56 0:29:09 0:30:25 0:31-28	0:29:27 0:30:15 0:28:50 0:28:50	0:27:54 0:29:56 0:25:16 0:25:16	0:25:09 0:29:21 0:29:02 0:31:59	- 0:28:45 $-$ 0:31:36	
	Off-peak Proportion (%)	0.7 25.3 8.0	3.4 18.7 0.8 18.7	19.5 19.5 16.3	13.0 13.0 0.8 16.2	0.3 20.5 20.5 20.5	44.1 - 17.8	
	eak ATT	$\begin{array}{c} 0.24.41 \\ 0.26.11 \\ 0.26.35 \\ 0.30.37 \end{array}$	0:25:16 0:25:45 0:27:58 0:27:8	0.26.47 0.27.20 0.27.03 0.27.03	0.2031 - 0.27:52 - 0.27:43	0.27.06 0.29.22 0.28.56 0.30.15	– 0:26:37 – 0:29:28	
	Evening p Proportion (%)	3.2 14.5 0.9	15.6 3.7 3.7	23.7 23.7 3.8 21.0		0.7 0.4 14.1	6.4 - 7 4	
	eak ATT	0.24:41 0.25:19 0.27:30 0.28:05	0:26:38 0:26:05 0:28:21 0:29:23	0:28:35 - 0.27:41	0:26:04 0:26:45 0:25:00 0:25:00		- $0.28:41$ $ 0:33:55$	
	Morning p Proportion (%)	14.5 17.7 2.8 7.1	5.6 2.6 2.6 2.6	0.2	13.5 14.1 18.8	8 - 1 8 - 1 9 - 1	16.6 - 7.6	
	Paths	2→© (FT) 2→© 6→© (FT) 6→©	© → © (FT) © → © (FT) © → © (FT)			€ (FT) (FT) (FT) (FT)	$\begin{array}{c} \overrightarrow{0} \rightarrow \overrightarrow{0}  (FT) \\ \overrightarrow{0} \rightarrow \overrightarrow{0}  (FT) \\ \overrightarrow{0} \rightarrow \overrightarrow{0}  (FT) \\ \overrightarrow{0} \rightarrow \overrightarrow{0}  (FT) \end{array}$	le by the authors
Table 7.         Detailed path choice         results	OD pairs	CWM→CGZX	CGZX→CWM	XD→CSS	CSS→XD	XS→BJXZ	BJXZ→XS	Source(s): Tabl

stations with little passenger flow. Virtual witnesses can be introduced to solve this problem in the subsequent research.

This study has resulted in an important finding for the Beijing Metro system. Frequent riders like commuters are better at choosing paths. Although the finding seems proving just common sense, this study enables an approach to uncovering quantified trip information with a high degree of certainty. Based on the finding of the path choices of frequent riders, thus specific while optimal paths can be recommended to other passengers. For individual passengers, if they do not know in advance what the waiting time will be, we can recommend optimal paths to them based on train schedules at time when they ride on the Beijing Metro system. In addition, when metro operation managers grasp the temporal and spatial distribution of passenger flow more accurately, they could prevent risks in areas with concentrated passenger flow in advance.

#### 6. Conclusions

We proposed a heuristic witness voting method by leveraging limited deterministic information on passengers' path choices to corroborate other uncertain information on their riding paths. By introducing the concept of witnesses, the path recovery method under simple rules that adapts to the complex metro networks is established. Passengers with the determinate path that a target passenger may encounter during the riding process are considered as witnesses; if witnesses at each phase of path cannot form a continuous and valid trip chain, it cannot be the target passenger's actual riding path, thus greatly reducing the number of possible paths. A voting mechanism was introduced to infer the most-likely riding path of the target passenger when there were multiple possible paths.

Since frequent riders like commuters are better at choosing paths based on our results, we can recommend better paths for other passengers by referring to the path choosing behaviors of frequent riders. Next, we can further combine individual passenger's walking speed if data becomes available and train schedules to recommend the path with the least possible riding time for him/her. Our future studies will also explore the implementation of these guidance strategies and an approach to quantifying the improved service quality and economic outcomes if fully implemented in real life.

#### References

- Bajaj, G., & Singh, P. (2021). Understanding preferences of Delhi metro users using choice-based conjoint analysis. *IEEE Transactions on Intelligent Transportation Systems*, 22(1), 384–393.
- Barry, J. J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in New York city with automated fare system data. *Transportation Research Record*, 1817, 183–187.
- Hörcher, D., Graham, D. J., & Anderson, R. J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological*, 95, 105–125.
- Hong, S. -P., Min, Y. -H., Park, M. -J., Kim, K. M., & Oh, S. M. (2016). Precise estimation of connections of metro passengers from Smart Card data. *Transportation*, 43(5), 749–769.
- Jin, F., Yao, E., Zhang, Y. and Liu, S. (2017). Metro passengers' route choice model and its application considering perceived transfer threshold. *PloS ONE*, Public Library of Science San Francisco, CA, 12(9), 1–17.
- Kusakabe, T., Iryo, T., & Asakura, Y. (2010). Estimation method for railway passengers' train choice behavior with smart card transaction data. *Transportation*, 37(5), 731–749.
- Li, W., Luo, Q., Cai, Q., & Zhang, X. F. (2018). Using smart card data trimmed by train schedule to analyze metro passenger route choice with synchronous clustering. *Journal of Advanced Transportation*, Hindawi, 2018, 2710608.1–2710608.13.

Megacity metro system

Liu, J. F., Sun, F.	. L., Bai, Y., & Xu, J. (2009)	. Passenger flow ro	ute assignment	model and algorithm	n for
urban rai	1 transit network. Journal	l of Transportation	ı Systems Engin	ueering and Informa	ation
Technolog	<i>gy</i> , <i>9</i> (2), 81–86.				

- Paul, E. C. (2010). Estimating train passenger load from automated data systems application to london underground. Cambridge, MA: Massachusetts Institute of Technology.
- Rashedi, Z., Mahmoud, M. E., Hasnine, S., & Habib, K. N. (2017). On the factors affecting the choice of regional transit for commuting in Greater Toronto and Hamilton Area: Application of an advanced RP-SP choice model. *Transportation Research Part A: Policy & Practice*, 105, 1–13.
- Raveau, S., Muñoz, J. C., & de Grange, L. (2011). A topological route choice model for metro. Transportation Research Part A: Policy and Practice, 45(2), 138–147.
- Si, B. F., Zhong, M., Liu, J. F., Gao, Z. Y., & Wu, J. J. (2013). Development of a transfer-cost-based logit assignment model for the Beijing rail transit network using automated fare collection data. *Journal of Advanced Transportation*, 47(3), 297–318.
- Su, G., Si, B., Zhao, F., & Li, H. (2022). Data-driven method for passenger path choice inference in congested subway network. Complexity. Hindawi, 2022, 5451017.
- Sun, Y., & Xu, R. (2012). Rail transit travel time reliability and estimation of passenger route choice behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2275(1), 58–67.
- Sun, L., Jin, J. G., Lee, D. -H., Axhausen, K. W., & Erath, A. (2014). Demand-driven timetable design for metro services. *Transportation Research Part C: Emerging Technologies*, 46, 284–299.
- Sun, L. J., Lu, Y., Jin, J. G., Lee, D. -H., & Axhausen, K. W. (2015). An integrated Bayesian approach for passenger flow assignment in metro networks. *Transportation Research Part C: Emerging Technologies*, 52, 116–131.
- Tang, L. Y., Zhao, Y., Cabrera, J., Ma, J., & Tsui, K. L. (2019). Forecasting short-term passenger flow: An empirical study on shenzhen metro. *IEEE Transactions on Intelligent Transportation Systems*, 20(10), 3613–3622.
- Van Veldhoven, Z., & Vanthienen, J. (2023). Best practices for digital transformation based on a systematic literature review. *Digital Transformation and Society*, 2(2), 104–128.
- Wang, L., Zhang, Y., Zhao, X., Liu, H., & Zhang, K. (2020). Irregular travel groups detection based on cascade clustering in urban subway. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 2216–2225.
- Wu, J., Qu, Y., Sun, H., Yin, H., Yan, X., & Zhao, J. (2019). Data-driven model for passenger route choice in urban metro network. *Physica A: Statistical Mechanics and Its Applications*, 524, 787–798.
- Xie, L. H., Zhang, Z. J., & Gong, D. Q. (2022). A heuristic close contact tracing method for urban rail transit. *Journal of Transportation Systems Engineering and Information Technology*, 22(4), 218–227.
- Xie, L. H., Zhang, Z. J., Qiu, R. G., Gong, D. Q., & Ma, Y. X. (2023). Heuristic witness voting to effectively identify passenger paths with limited deterministic information in an urban rail transit system. *IEEE Transactions on Intelligent Transportation Systems*, submitted for publication.
- Xu, R. H., Luo, Q., & Gao, P. (2009). Passenger flow distribution model and algorithm for urban rail transit network based on multi-route choice. *Journal of the China Railway Society*, 31(2), 110–114.
- Xu, X. Y., Xie, L. P., Li, H. Y., & Qin, L. Q. (2018). Learning the route choice behavior of subway passengers from AFC data. *Expert Systems with Applications*, 95, 324–332.
- Zhang, Y. S., Yao, E. J., Zhang, J. Y., & Zheng, K. N. (2018). Estimating metro passengers' path choices by combining self-reported revealed preference and smart card data. *Transportation Research Part C: Emerging Technologies*, 92, 76–89.

DTS

2.3

- Zhao, J., Qu, Q., Zhang, F., Xu, C., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems*, 18(11), 3135–3146.
- Zhao, J., Zhang, F., Tu, L., Xu, C., Shen, D., Tian, C., et al. (2017). Estimation of passenger route choice pattern using smart card data for complex metro systems. *IEEE Transactions on Intelligent Transportation Systems*, 18(4), 790–801.
- Zhou, F., & Xu, R. H. (2012). Model of passenger flow assignment for urban rail transit based on entry and exit time constraints. *Transportation Research Record*, 2284(1), 57–61.
- Zhou, F., Shi, J., & Xu, R. (2015). Estimation method of path-selecting proportion for urban rail transit based on AFC data, edited by Fan, W. *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, Vol. 2015, p. 350397.
- Zhu, W., Hu, H., & Huang, Z. D. (2014). Calibrating rail transit assignment models with genetic algorithm and automated fare collection data. *Computer-Aided Civil and Infrastructure Engineering*, 29(7), 518–530.
- Zhu, Y. W., Koutsopoulos, H. N., & Wilson, N. H. M. (2021). Passenger itinerary inference model for congested urban rail networks. *Transportation Research Part C: Emerging Technologies*, 123, 102896.

### Further reading

- Bovy, P. H. L., & Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. *Transportation*, 32(4), 341–368.
- Daganzo, C. F., & Sheffi, Y. (1977). On stochastic models of traffic assignment. Transportation Science, 11(3), 253–274.
- Zhu, Y. W., Koutsopoulos, H. N., & Wilson, N. H. M. (2017). Inferring left behind passengers in congested metro systems from automated data. *Transportation Research Procedia*, 23, 362–379.

#### **Corresponding author**

Daqing Gong can be contacted at: dqgong@bjtu.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm Or contact us for further details: permissions@emeraldinsight.com 339

Megacity

metro system