

Received November 21, 2019, accepted January 2, 2020, date of publication January 9, 2020, date of current version January 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2965155

Estimate Passengers' Walking and Waiting Time in Metro Station Using Smart Card Data (SCD)

WANJUN LI¹, XUEDONG YAN¹, XIAOMENG LI², AND JINGSI YANG¹

¹MOT Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

²Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology, Kelvin Grove, QLD 4059, Australia

Corresponding author: Xuedong Yan (xdyan@bjtu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2019JBZ003.

ABSTRACT Passengers' walking time and waiting time at metro stations are important indicators for evaluating the level of service (LOS) of a metro station. This paper aims to estimate passengers' walking time and waiting time at metro stations based on a proposed passenger-to-train assignment method using smart card data (SCD). Firstly, two algorithms were developed to determine the possible train choice set for all the passengers. Secondly, passengers' walking time distribution was estimated for different group of passengers categorized by travel periods, type of cardholder and social demographics. Lastly, passengers' waiting time at origin stations was explored based on the train choice model proposed in this study. The study found that passengers traveling at peak hours tend to walk faster than they do at non-peak hours. The overall waiting time at origin stations is highly related to the train headway (time interval between the adjacent arrival trains in the same direction and station) and passenger volume. More importantly, the result suggested when the passenger volume is low, train headway could significantly affect waiting time. However, when the passenger volume exceeds certain level (>1000 in 15 minutes), it becomes to the major factor leading to the increase in waiting time. Elders and disabled passengers have extended waiting time during peak hours than other passengers. The method proposed in this study offers a new perspective of using SCD in travel behavior analysis. Besides, the results bring insights for future metro station construction and operation design as well as evaluation on the LOS of metro system.

INDEX TERMS Level of service (LOS), metro system, smart card data (SCD), waiting time, walking time.

I. INTRODUCTION

Metro system, with high efficiency and sustainability, is one of the most popular public transportation modes widely developed in many cities in China. In comparison to other public transportation systems, metro system provides more reliable travel time due to its highly-organized and independent operation system. In Beijing, the daily maximum passenger volume of metro system can reach up to ten million since 2018 [1]. Passengers' experience in metro system is closely related to architectural structure design and the system's level of service (LOS). To enhance the passengers' satisfaction, metro system designers and operators should focus on improving the service level from different perspectives.

The associate editor coordinating the review of this manuscript and approving it for publication was Roberto Sacile.

Passengers' waiting time at origin station is an important indicator that measures the operation performance of the metro system. Previous studies have pointed out that passengers are more likely to perceive waiting time compared to in-vehicle time [2-4]. Waiting time is related to several factors, such as station environment, time-of-day, and travel purpose. It has been reported that random passengers (who are not arranging their arrival time) are more likely to arrive at non-peak hours, but passengers traveling at peak hours tend to experience shorter waiting time [5-6]. Recent study found that station auxiliary facilities, such as shelters and benches, have impact on passengers' perception of waiting time [3].

Though waiting time plays an important role in LOS evaluation, challenges remain regarding how to obtain waiting time practically. Many public transit assignment models, especially the frequency-based models usually assume that passengers arrive randomly and the average waiting time

equals half of headway in both bus and metro systems [7–13]. Some studies investigating passengers' arrival patterns also applied simple linear relationship between waiting time and headway [14]. However, other studies believed that passengers should not be treated with random arrival as a whole [15]. With the open access to public transit timetable, threshold has been explored between arranged arrival of passengers (who arrange their arrival time based on train schedule to minimize their waiting time) and random arrival of passengers. Some researchers suggest that 10 to 11 minutes headway is the dividing line between random passengers and non-random passengers, and passengers tend to arrange their arrival time to minimize waiting time when the headway exceeds 38 min [16]–[18].

The conventional methods to directly obtain passenger's waiting time consist of field observation and personal survey. In recent decades, the widespread use of smart card for public transport brings new perspective for waiting time estimation. Smart card data (SCD), which collect travel information at both entry and exit gates of metro system, minimizes the human role in data collection [19]. Ingvarsson et al. developed a method to estimate waiting time using smart card data by assigning the traveler to the first train running between the origin and destination stations [20]. Specifically, the study counted waiting time as the period from tap-in to the departure time of assigned train, but this situation was not exactly true, since passengers' walking time from the entry gate to the platform was included in the waiting time.

Walking time, as the time tapping the card to enter the station to the platform, is different period from waiting time, is another important indicator that should be considered throughout every trip from entry to exit. For passengers, walking time to get access to the train directly influences their train choices and waiting time on the platform. For metro system operators, walking time reflects the rationality of station design and provides valuable information when it comes to improve the system LOS.

Previous studies have applied different methods to estimate passengers' walking time. Some studies assumed that passengers with the shortest travel time of each Origin-Destination (OD) pair have zero waiting time and applied regression model to estimate the total walking time in the trip [21]–[22]. However, how to define the “shortest” travel time seems to be a practical issue. Therefore, field survey has been commonly used to collect walking time information. Sun and Xu selected eight stations with a sample size of 40 passengers to determine the proper distribution of walking time [23]. Chen et al. conducted passenger travel trajectory analysis and estimated the walking time by field survey [24]. Similarly, some studies explored the route choice behavior or passenger flow assignment by using fixed walking time determined by survey data [25]–[26]. In addition, some studies obtained the walking time using walking distance divided by walking speed. For example, Zhu et al. proposed a method by using walking distance obtained from station plan divided by the average walking speed of 1.12 m/s (derived from agency

observation) to obtain the exit time distribution [27]–[28]. Zou et al. calculated the access and egress walking time in corresponding to the walk distance and the average walking speed [29].

The traditional methods resorting to surveys or observations to obtain passengers' walking time and waiting time are usually time-consuming, which requires high cost on human resources. With the availability of smart card data (SCD), this study provides a new approach to estimate passengers' walking and waiting time. Specifically, the study has four objectives: 1) to deduce passengers' train choice set using smart card data and train schedule timetable; 2) to estimate passengers' walking time and waiting time in a given metro station; 3) to explore the walking time distribution among travel time-of-day (peak vs. non-peak hours), passenger characteristics (elder & disabled vs. normal passengers) and cardholder type (long-term vs. one-time); and 4) to explore the relationship among train headway, passenger volume and waiting time. Following the research objectives, several assumptions need to be clarified prior to proceeding the study. These assumptions are: 1) all passengers arrive at the station randomly; 2) passengers entering a station share the same walking path as passengers exiting the station; 3) all trains run punctually in accordance with the timetable; 4) passengers walk out of the destination station without lingering after they get off the train; 5) capacity constraint is not considered when trains are assigned.

In the following Section 2, the concept of travel trajectory in metro system and definition of passengers with single feasible train choice are introduced. Section 3 presents the algorithms design on passenger-to-train assignment, which aims to determine the possible train choice set for all passengers. Meanwhile, the methodology of walking time estimation, train selection and waiting time estimation is introduced in this section. Estimation results of four selected stations in Nanjing Metro system for walking time, waiting time and passenger delay are presented in Section 4. Discussion with practical application are provided in Section 5, and conclusions are drawn in Section 6.

II. COCEPTS IN THE STUDY

A. TAVEL TRAJECTORY AND TIME PERIODS

The determination of travel trajectory in metro system is relatively simpler than that in roadway system, especially with the information provided by smart card data (SCD). A metro station can be divided into two main areas, i.e. pay zone and free zone. The information recorded by SCD only covers the time period within pay zone. When passengers swipe the smart card at the station entry or exit gate, the station ID and entry/exit time are recorded. In general, travel time without transfer contains four parts, which are (1) access walking time (AWT, i.e. time from the entry gate to the platform), (2) platform waiting time (PWT), (3) in-train running time (TRT), (4) egress walking time (EWT, i.e. time from passengers off the train to the exit gate). Figure 1 shows a

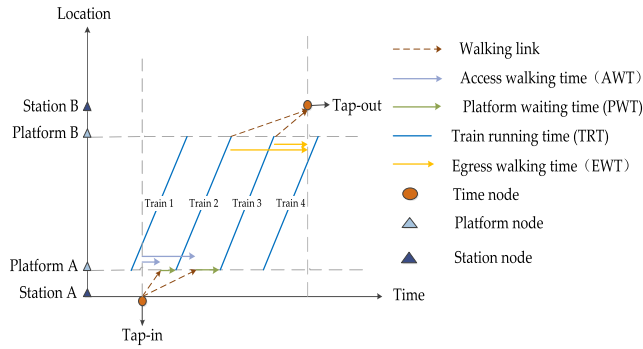


FIGURE 1. Spatiotemporal travel trajectory without transfer.

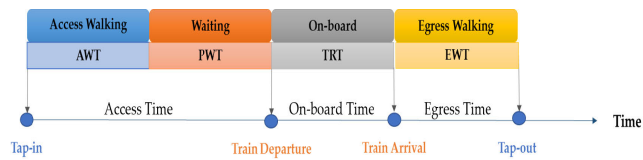


FIGURE 2. Time composition of traveling.

spatiotemporal travel process of passenger from entry to exit without transfer. In this particular situation, two trains (No.2 and No.3) are operated between the time interval from tap-in to tap-out.

Since the time-related information within station is limited from SCD, it is difficult to determine the possible train choice of passengers. However, if combining the operation train timetable with SCD, more information can be extracted. Figure 2 illustrates the time periods of traveling in metro system. Four types of activities occur in four periods, and the certain time nodes can be determined once the boarding train is confirmed.

B. PASSENGERS WITH SINGLE FEASIBLE TRAIN CHOICE

The travel time between the same origin-destination (OD) stations could be very different among different passengers. Figure 3 describes two common situations of traveling between an OD pair. In situation 1, the passenger has only one feasible train choice within the travel period. However, in situation 2, as the travel time extends, more choices become available. Passengers with single feasible train choice can be recognized as a group of people having minimum waiting time and optimized travel time between any OD pairs. If we assume that the passengers walk out of the stations without lingering (*assumption 4*), which is mostly true in the real life, and combine it with *assumption 2*, the egress walking time can then represent the access walking time of the same station. With the availability of access walking time, passengers' waiting time on the platform can be obtained.

III. METHODOLOGY

The methodology consists of three parts. The first part (3.1&3.2) describes the two algorithms developed to form train choice set for all the passengers. Meanwhile, the

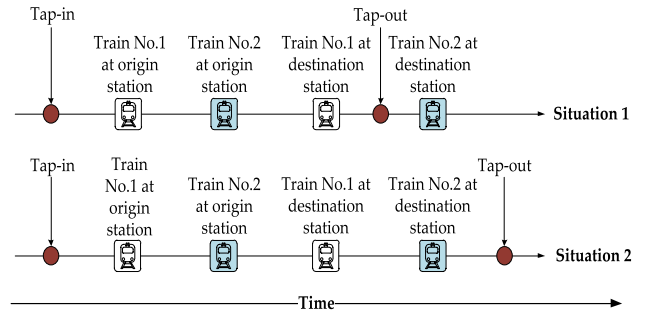


FIGURE 3. Train choice situations in time sequence.

passengers with single feasible train choice are derived from the algorithms and used as data input in the second part. The second part (3.3) introduces the application of Maximum Likelihood Method (MLE) on station walking time distribution estimation. The last part (3.4) performs train assignment method for passengers who have multiple train options, and estimates the overall waiting time based on train assignment result.

A. NOTATIONS

The notations and parameters applied in passenger-to-train algorithms are listed in Table 1.

Parameters calculation equations:

$$c_{s_i, s_j}^{l_r} = t_h^{\text{arrival}, s_j, l_r} - t_h^{\text{arrival}, s_i, l_r} \quad (1)$$

$$c_{\text{access}, p, h}^{s_o} = t_h^{\text{arrival}, s_o, l_r} - t_{\text{in}, p}^{s_o} \quad (2)$$

$$c_{\text{egress}, p, h}^{s_d} = t_{\text{out}, p}^{s_d} - t_h^{\text{arrival}, s_d, l_r} \quad (3)$$

$$c_p^{s_o, s_d} = t_{\text{out}, p}^{s_d} - t_{\text{in}, p}^{s_o} \quad (4)$$

B. PASSENGER-TO-TRAIN ALGORITHMS

To identify the possible train set for passengers, two algorithms are developed based on different assumptions and constraints.

The first algorithm is established based on the assumption that passengers would board the first coming train to minimize the access time at origin station without considering train capacity constraint (*equation 5*).

$$t_h^{\text{arrival}, s_o, l_r} - t_{\text{in}, p}^{s_o} \geq 0, \quad t_h^{\text{arrival}, s_o, l_r}, t_{\text{in}, p}^{s_o} \in T, \quad l_r \in L, s_o \in S, \quad (5)$$

As such, the selected train should satisfy the constraint of *equation 6*, which indicates that the egress time of passengers on the assigned train should not be less than zero.

$$c_{\text{egress}, p, h}^{s_d} \geq 0, \quad s_d \in S, h \in H, p \in P, \quad (6)$$

If the constraint is satisfied, the train will be selected and included in the set H^{S_1} .

The second algorithm is established based on the assumption that passengers would exit the destination stations without lingering (*equation 7*). They are assigned to the last

TABLE 1. Notations and parameters.

Notation	Definition
S	Set of all stations
O	Set of origin stations,
D	Set of destination stations, $D \subset S$
T	Set of time node
L	Set of lines connected by stations
H	Set of scheduled trains in the system
H^{S_1}	Set of selected trains in the algorithm 1
H^{S_2}	Set of selected trains in the algorithm 2
P	Set of passengers
h	Index of assigned train, $h = 1, 2, \dots, n, h \in H$
$l_{r,o,d}$	Index of lines for route r connecting an OD pair (o, d), $h = 1, 2, \dots, n, h \in H$
s_i, s_j	Index of stations, $s_i, s_j \in S$
s_o	Index of origin stations, $s_o \in S$
s_d	Index of destination stations, $s_d \in S$
p	Index of passenger, $p \in P$
$t_h^{\text{arrival}, s_i, l_r}$	Index of time node of train h arrival at station s_i on line l_r , $t_h^{\text{arrival}, s_i, l_r} \in T, s_i \in S, l_r \in L, h \in H$
$c_{s_i, s_j}^{l_r}$	Train running time cost between station i and j on line l_r , $l_r \in L, s_i, s_j \in S, h \in H$
$t_{\text{in}, p}^{s_o}$	Index of time node recorded on smart card data of passenger p entering the origin station s_o , $t_{\text{in}, p}^{s_o} \in T, s_o \in S, p \in P$
$t_{\text{out}, p}^{s_d}$	Index of time node recorded on smart card of passenger p exiting from the destination station, $s_d, t_{\text{out}, p}^{s_d} \in T, s_d \in S, p \in P$
$c_{\text{access}, p, h}^{s_o}$	Time cost of passenger p from accessing the origin station till boarding the selected train h , $s_o \in S, p \in P, h \in H$
$c_{\text{egress}, p, h}^{s_d}$	Time cost of passenger p egressing the destination station after get off from the selected train h , $s_d \in S, p \in P, h \in H$
$c_{\text{access}, p, h}^{s_o, \text{walking}}$	Time cost of passenger p walking from tap-in to the platform at origin station, $s_o \in S, p \in P, h \in H$
$c_{\text{access}, p, h}^{s_o, \text{waiting}}$	Time cost of passenger p from arriving the platform till boarding the selected train h at origin station, $s_o \in S, p \in P, h \in H$
$c_p^{s_o, s_d}$	Time cost from tap-in to tap-out of passenger p , $s_d, s_o \in S, p \in P$

arrival train before tapping-out at destination station without considering train capacity constraint.

$$t_{\text{out}, p}^{s_d} - t_h^{\text{arrival}, s_o, l_r} \geq 0, \quad t_{\text{out}, p}^{s_d}, t_h^{\text{arrival}, s_o, l_r} \in T, \quad l_r \in L, s_d \in S, h \in H, p \in P \quad (7)$$

In the same manner, the selected train should satisfy the constraint that the access time should not be less than zero (equation 8).

$$c_{\text{access}, p, h}^{s_o} \geq 0, \quad s_o \in S, h \in H, p \in P \quad (8)$$

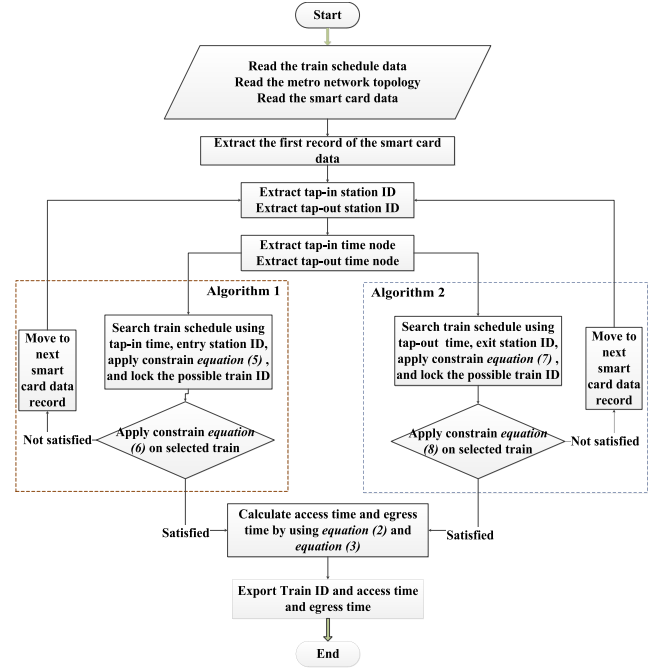
If the constraint is satisfied, the train will be selected and included in the set H^{S_2} .

Through the two algorithms, the assigned train number, passenger access time and egress time of selected train can be obtained (equation 2 & equation 3).

The flowchart of the algorithms is shown in Figure 4.

C. ESTIMATION OF WALKING TIME DISTRIBUTION

Passenger train choice set can be deduced from the two developed algorithms. The first algorithm determined the first train

**FIGURE 4.** Passenger-to-train assignment algorithms framework.

for passengers within the travel time while the second algorithm deduced the last possible train. By combining the two algorithms, a group of passengers with the same train choice can be identified, namely, *passengers with single feasible train choice*. Their egress walking time provides valuable information to estimate the overall average walking time by applying *assumption 2*.

For the same walking distance, a lognormal distribution is adopted for egress walking time since the walking speed at metro station is usually considered as normal or log-normal distribution [30]–[33]. To ensure an unbiased and asymptotically efficient estimation on population mean and variance of walking time, the Maximum Likelihood Estimation (MLE) is applied in this study.

Maximum likelihood function:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma x_i} e^{-\frac{(\ln x_i - \mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} (\prod_{i=1}^n x_i)^{-1} \\ &\quad \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2\right), \\ &\quad x_i > 0 (i = 1, 2, \dots, n) \end{aligned} \quad (9)$$

Take logarithm for maximum likelihood function:

$$\begin{aligned} \ln L(\mu, \sigma^2) &= -\frac{n}{2} \ln(2\pi\sigma^2) - \ln \prod_{i=1}^n x_i \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2 \end{aligned} \quad (10)$$

Use derivation method for parameter μ and σ^2 :

$$\left\{ \begin{array}{l} \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (\ln x_i - \mu) = 0 \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\ln x_i - \mu)^2 = 0 \end{array} \right\} \quad (11)$$

Solve equation (11), and get (12) and (13),

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (12)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \frac{1}{n} \sum_{i=1}^n \ln x_i)^2 \quad (13)$$

The mean and variance of lognormal distribution are,

$$EX = e^{\hat{\mu} + \frac{\hat{\sigma}^2}{2}} \quad (14)$$

$$DX = e^{2\hat{\mu} + \hat{\sigma}^2} (e^{\hat{\sigma}^2} - 1) \quad (15)$$

D. TRAIN ASSIGNMENT AND WAITING TIME ESTIMATION

In addition to passengers with single feasible train choice, some passengers have more than one train options within their travel periods. The principle of assigning these passengers to the most possible train is based on the assumption that passengers walk out of the destination station as soon as possible without lingering (*assumption 4*). In such situation, their egress walking time should be closest to the estimated station mean walking time. Thus, the study used the Probability Density Function (PDF) (*equation 16*) on egress walking time for different train choices, where x is the egress walking time of the selected train, μ is the mean walking time of destination station, and σ^2 is the variance. The function determines the possibility of egress time x following the lognormal distribution of $\ln X \sim N(\mu, \sigma^2)$. The train will be selected when the egress walking time has the highest estimation value in *equation 16*.

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (16)$$

While passengers are assigned to the most possible train, their overall access time can be calculated. Referring to 3.3, the average walking time has been obtained. Therefore, passengers' waiting time can be calculated by the following *equation 17*.

$$c_{access,p,h}^{s_o,waiting} = c_{access,p,h}^{s_o} - \bar{c}_{access,p,h}^{s_o,walking} \quad (17)$$

where, $\bar{c}_{access,p,h}^{s_o,walking}$ represents the mean walking time of each station obtained by MLE.

In order to evaluate the average waiting time in different time periods, the entry time recorded in smart card data is sliced by every 15 minutes from 6:00 to 24:00. The average waiting time is calculated by *equation 18*, where \bar{c} is the estimated mean walking time and N_p is the total number of entered passengers in a 15-min time slice.

$$w = \frac{\sum_t' (c_{access,p,h}^{t,t'} - \bar{c})}{N_p} \quad (18)$$



FIGURE 5. Nanjing metro topology network in 2014.

TABLE 2. Trips generated and terminated at selected stations.

Category	Station ID	Entry (1)	Entry (>1)	Exit (1)	Exit (>1)	Total
Small-scale	31	3807	3083	4370	3004	14264
	3	1382	776	1388	676	4222
Large-scale	14	22117	15900	19662	14334	72013
	16	23312	21255	23750	14839	83156

IV. RESULTS

Four stations are selected from Nanjing Metro system to perform the overall estimation results. They are categorized into small-scale and large-scale stations which is determined by the total passenger volume of entry and exit in one workday. The results are presented in four sections: (1) a brief description of data input; (2) walking time estimation results of four selected stations; (3) waiting time estimation results of four selected stations; (4) passengers' total delay estimation at the origin stations.

A. DATA DESCRIPTION

The data used in this study were provided by Nanjing (China) Metro Group Company, Ltd. in 2014. Figure 5 shows a simplified Nanjing Metro network in 2014. It consisted of five lines in total and one of them (S8) is disconnected from the other four lines. There are 88 stations on these lines, between which over 1500 trains are operated every day. A typical workday data of four stations were used in this study (November, 03, 2014) and the trips with transfer were excluded.

By applying the algorithms proposed in the above section, the possible train sets for each trip are determined. Data descriptions of selected stations are listed in Table 2. In specific, station 14 is close to Nanjing train station and station 31 is nearing one of the most famous attraction spots in Nanjing. Trips generated and terminated at each station are composed by two types of passengers: passengers with single feasible train choice (1) and passengers with more than one train choices within their travel period (>1).

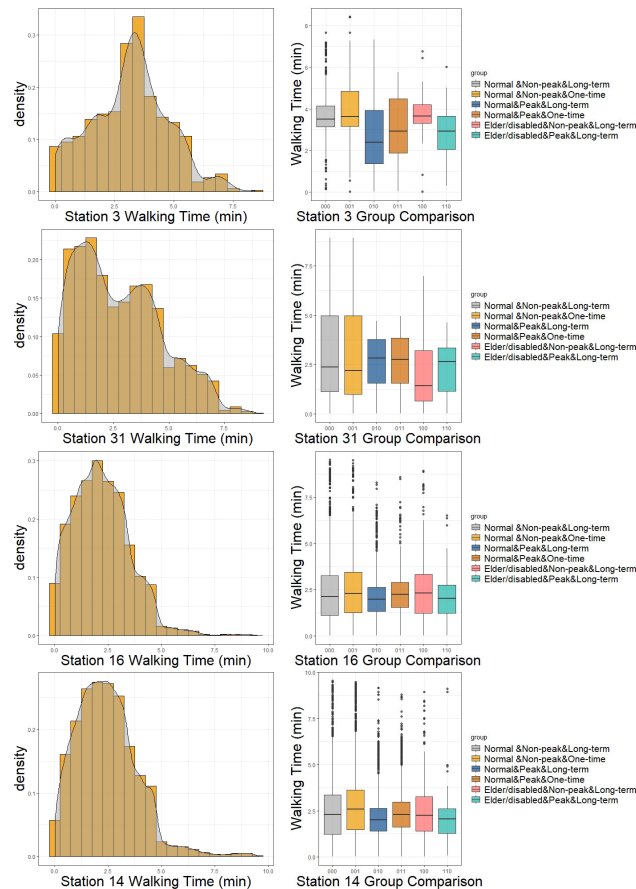


FIGURE 6. Walking time distribution and box plot of different groups of passengers.

B. PASSENGERS' WALKING TIME DISTRIBUTION

The walking time estimation was based on passengers with single feasible train choice. Three potential factors relevant to passengers' walking time distribution in different size of stations are explored, namely, passenger characteristics (elder/disabled passengers vs. normal passengers), travel period (non-peak vs. peak), and cardholder type (long-term vs. one-time). Specifically, the long-term card represents stored-value card and one-time card represents single-use recycled card.

The walking time distribution of four stations and the box plot of different groups of passengers are shown in Figure 6. The estimation results suggest that most passengers spend two to five minutes in walking. The long-tail in large-scale stations indicates that there is a small group of passengers walking extremely slowly compared to the others.

There is noticeable passenger group difference in average walking time in small-scale stations, whereas the group difference in large-scale stations is very small. In most stations, the normal passenger group who hold long-term card and travel in peak hours have the shortest average walking time. Moreover, elder/disabled passengers and passengers with one-time card traveling in non-peak hours have longer walking time than other groups.

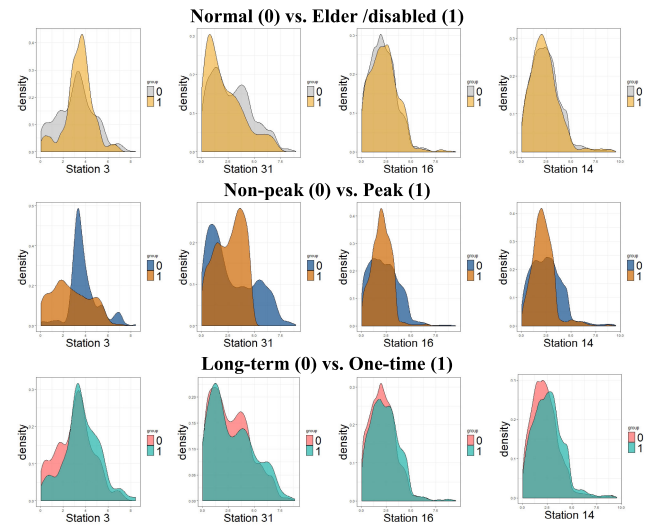


FIGURE 7. Walking time distribution comparison in different groups.

In order to explore the variance within and between groups, the walking time distribution comparison are plotted in Figure 7, from which three major tendencies can be observed. Firstly, elder/disabled passengers mostly walk in an average-level speed. Secondly, the walking time distribution during peak hours is more concentrated than that in non-peak hours. Lastly, passengers holding one-time cards tend to have longer walking time than other passengers holding long-term cards.

C. PASSENGERS' WAITING TIME

Waiting time estimation is based on the train assignment method proposed in the previous section. The average waiting time for all passengers and elder/disabled passengers are presented in Figure 8. The result shows that the duration of waiting time is greatly affected by train headway and passenger volume.

In the most situations, the average waiting time decreases with the reduction of train headway. However, when the headway exceeds 6 minutes, the overall waiting time increases remarkably. Besides, large passenger volume also contributes to long waiting time even when the headway is small. For example, in Station 14, when the passenger volume is over 1500 within 15 minutes, the average waiting time increases to 2 minutes which exceeds half of the train headway. Station 16 shows a worse situation when passenger volume is over 3000 within 15 minutes, as the average waiting time is longer than headway. This means a part of passengers are not able to board the first coming train due to over crowdedness.

As shown in Figure 8, elder/disabled passengers are more likely to wait longer time during peak hours. This is probably because elder/disabled passengers are more willing to extend their waiting time to increase the chance of having a seat than normal passengers.

Moreover, the figure 8 demonstrates that headway has strong influence on average waiting time especially when

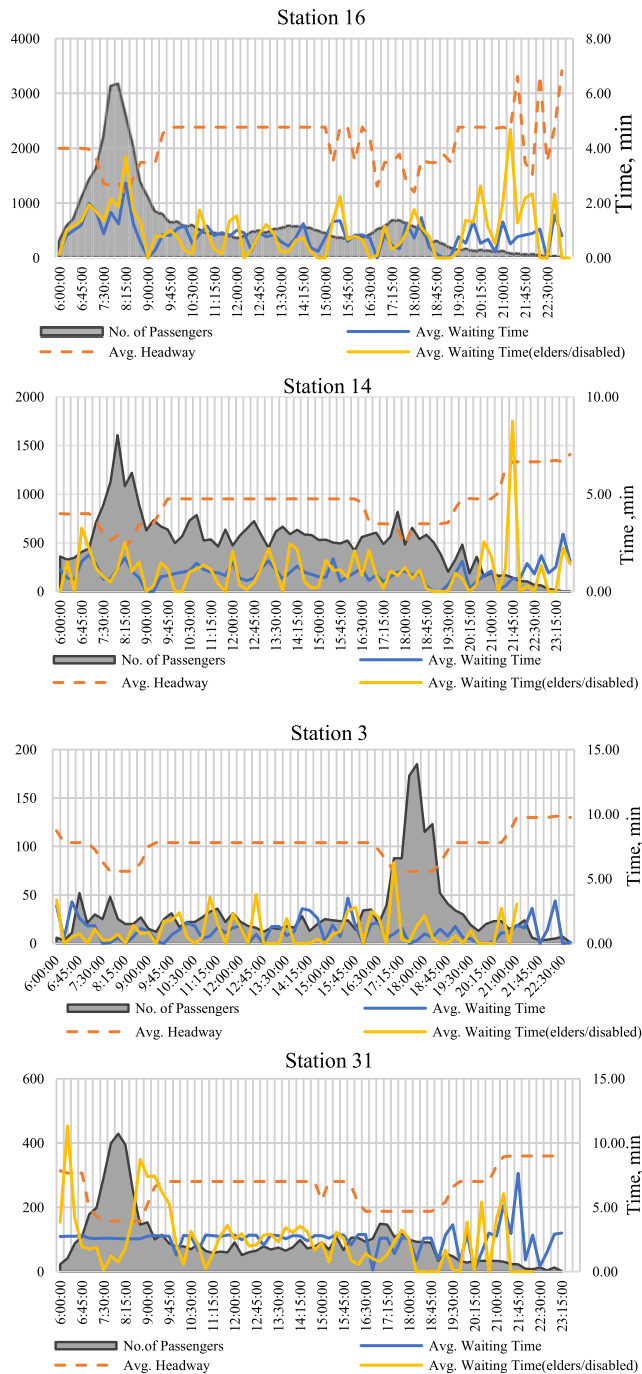


FIGURE 8. Average waiting time estimation result in all day.

the passenger volume is low. However, as the passenger volume reaches a critical point, shortening train headway has no substantial impact on reducing the average waiting time. To explore this critical point, the relationship between passenger volume and average waiting time at four stations is explored and presented in Figure 9. The result suggests when the number of entering passengers is over 1000 in 15 minutes, the waiting time starts to increase even though the train headway is small.

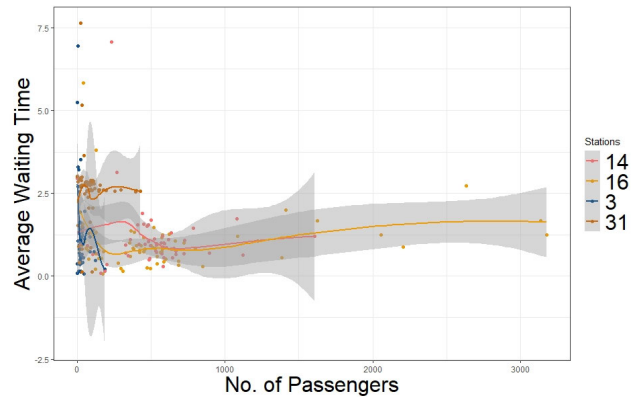


FIGURE 9. Relationship between average Waiting time and passenger volume.

TABLE 3. Mann-Whitney U test for different groups of passengers on total access time (delay).

Group	Mean(min)	Mann-Whitney U Test Sig.
Normal vs. Elder/disabled	Normal (3.62) Elder/disabled (3.86)	0.000
Non-peak vs. Peak	Non-peak (3.72) Peak (3.51)	0.000
Long-term vs. One-time	Long-term (3.56) One-time (3.76)	0.000
Small-scale vs. Large-scale	Small-scale (4.57) Large-scale (3.53)	0.000

*Significance level= 0.05

D. PASSENGERS' DELAY AT ORIGIN STATIONS

The LOS of metro system should be enhanced through reducing passengers' walking distance in the station and their waiting time on the platform. In this study, the total time of walking and waiting (access time) in metro station is used as passengers' delay to get access to the train. To examine the differences of passenger delay between different passenger groups and stations, the Mann-Whitney U test is adopted since the data have a non-normal distribution. The results are listed in Table 3. According to the results, passenger characteristics, travel period, cardholder type and station scale all have significant impact on passengers' delay. Elder/disabled passengers have significant longer delay than normal passengers. People travelling at peak hours have shorter delay than those travelling at non-peak hours, and long-term card users have shorter delay than one-time card users. Moreover, large-scale stations are associated with significantly shorter delay compared with small-scale stations.

V. DISCUSSIONS

Smart cards used in metro system contain substantial amount of data, and relatively accurate information of passengers' spatiotemporal movement within a metro system. The algorithms developed in this study make full use of the SCD and train scheduled timetable to provide valuable information (e.g. passenger walking time and waiting time) for the metro system LOS evaluation. The algorithms are developed

TABLE 4. Passenger-to-train methods comparison.

	Core Concept	Advantages	Disadvantages
Zhu, 2017 [27-28]	Probabilistic method (Bayes' theorem)	Consider the train capacity constrain	Require field observation data input in the model
		The model can be validated	Passengers with transfer are not included
This study	Data-driven inferred algorithm	Field survey data input is not required	Apply rigid assumption on passenger travel behavior
		Investigate differences between different types of passengers	Passengers with transfer are not included

based on several reasonable assumptions which have also been adopted in previous research [34-36]. Table 4 gives the comparison between passenger-to-train methods proposed by different studies.

Previous studies mentioned that passengers with single feasible train choice can only represent the passengers with faster walking speed [27-28]. In this study, passengers with single feasible train choice account for about half of the travelers, which is a considerable amount. Moreover, the average walking time deduced from this group of passengers is around two to five minutes, which is consistent with the estimated walking time in Beijing Metro Network [24]. This result suggests that passengers with single feasible train choice benefit from short waiting time on the platform, rather than fast walking speed. Another study conducted in Singapore reported a quite short walking time compared to the present study, which is 20 to 70 seconds [22]. It is suspected that the big difference in walking time largely depends on the station design in different cities/countries, instead of passengers' walking speed.

Passengers with different characteristics, holding different types of cards and travelling at different periods and stations present different walking patterns. During peak hours, passengers tend to walk at a faster speed than they do at non-peak hours. This can be explained as during peak hours, the travelers are mostly commuters who are familiar with the path and have strict time limit, and a short walking time can decrease the overall travel time to some extent.

Elder/disabled passengers travelling at non-peak hours were the group with longest walking time. Passengers holding one-time recycling cards walk more slowly than long-term card holders, probably due to their unfamiliarity of the path. Large-scale stations undertake more trips compared to small-scale stations, and thus the differences between passenger groups are reduced in large-scale stations.

The estimation result of waiting time suggests that train headway is a critical factor, and large headway can greatly extend the overall waiting time. In addition, the increase

passenger volume during peak hours could increase the overall waiting time but the increment is not remarkable. The possible reason relates to a frequent train arrival in peak hours which somehow offsets the increasing tendency, and a similar finding has been reported by Sun and Xu [23]. However, a critical threshold of entering passenger volume (1000 in 15 minutes) was found, over which the waiting time starts to increase even though the train headway is short.

The overall access time shows discrepancy between different groups of passengers, and specific group of passengers have longer delay than the others. For example, elder/disabled passengers and passengers holding one-time card are more likely to spend longer time at origin stations. Furthermore, passengers travelling at peak hours and large-scale stations have less delayed than those passengers travelling in non-peak hours and small-scaled stations.

Overall, results of the study have several practical implications on metro station design, train operation and passenger travelling plan. Walking time is an important measurement that could be used in evaluating the rationality of a station design. The study suggests that 2-5 min is an average range for walking time, but it also depends on the passenger features and the station design. The walking time estimation result in this study provides conclusive evidence that some stations in Nanjing metro system require improvement on access facilities and guide sign to reduce access time for special groups of passengers.

Passengers' waiting time should be managed by considering both the headway and the passenger volume. Meanwhile, as a vulnerable group, more attention should be paid on elder/disabled passengers, since they have longer walking time and waiting time than the other passengers, especially during the peak hours.

Some limitations should be mentioned in this study. Firstly, one should note that there is always a minor gap between scheduled timetable and actual operation time. Using scheduled timetable may cause a bias in the estimation results. The real operation timetable (if available) should be applied in future studies to optimize the results. Secondly, the number of SCD inputs can be enlarged. Larger data input can improve the estimation accuracy and decrease the variance within each group. Third, the walking time distribution captured in this study only represents the average speed level. The passengers with single feasible train choice somehow exclude people with slower walking speed. Moreover, the walking time in this study only consider the distance from entry gate to the platform due to the lack of information in SCD. Lastly, the assumption of neglecting train capacity constraint should be modified in the future study.

VI. CONCLUSION

In summary, this study developed two algorithms to determine the train choice set for all the passengers. The algorithms identified passengers with only one possible train choice within travel period, and provided dependable estimation on station walking time and waiting time. With detailed

information provided by SCD, the differences in walking time in relation to travel periods and personal characteristics were investigated. Moreover, based on the walking time estimation, a new train assignment method was proposed in this study. The assignment results provided explicit travel information of all the passengers, including their waiting time on the platform. The result revealed the effect of passenger volume and headway design on overall waiting time. Further, it demonstrated the travel behavior difference related to passenger attributes and travel period.

The application of SCD in this study brings a new perspective for the research of passenger-to-train assignment analysis. Findings of the study provide implications for practical evaluation of operation and management performance of metro system, and offer referential information on timetable optimization and metro station design. Future study will focus on developing integrated train assignment method in metro network by considering the transferring passengers.

CONFLICTS OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

REFERENCES

- [1] Z. J. Chen, "Real-time estimation model and control method for urban rail transit line load," Ph.D. dissertation, School Traffic Transp., Beijing Jiaotong Univ., Beijing, China, 2018.
- [2] O. A. Nielsen, "A stochastic transit assignment model considering differences in passengers utility functions," *Transp. Res. B, Methodol.*, vol. 34, no. 5, pp. 377–402, Jun. 2000.
- [3] Y. Fan, A. Guthrie, and D. Levinson, "Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security," *Transp. Res. A, Policy Pract.*, vol. 88, pp. 251–264, Jun. 2016, doi: 10.1016/j.tra.2016.04.012.
- [4] M. Fosgerau, K. Hjorth, and S. L.-J. Vincent, *The Danish Value of Time Study: Results for Experiment 2*. Copenhagen, Denmark: The Danish Transport Research Institute, 2007.
- [5] G. Currie and D. R. Csikos, "The impacts of transit reliability on wait time: Insights from automated fare collection system data," Presented at the Transp. Res. Board 86th Annu. Meeting, p. nos. 07–0544, 2007. [Online]. Available: <https://trid.trb.org/view/801165>
- [6] M. Luethi, U. A. Weidmann, and A. Nash, "Passenger arrival rates at public transport stations," Presented at the Transp. Res. Board 86th Annual. Meet., 2007. [Online]. Available: <https://trid.trb.org/view.aspx?id=801220>
- [7] S. Hutchinson, "Waiting for the bus," *Social Text*, vol. 18, no. 2, pp. 107–120, 2000.
- [8] Y. Liu, J. Bunker, and L. Ferreira, "Transit users' route-choice modelling in transit assignment: A review," *Transp. Rev.*, vol. 30, no. 6, pp. 753–769, Nov. 2010.
- [9] Q. Fu, R. Liu, and S. Hess, "A review on transit assignment modelling approaches to congested networks: A new perspective," *Procedia Social Behav. Sci.*, vol. 54, pp. 1145–1155, Oct. 2012.
- [10] K. Nökel and S. Weckel, "Boarding and alighting in frequency-based transit assignment," *Transp. Res. Rec.*, vol. 2111, no. 1, pp. 60–67, Jan. 2009, doi: 10.3141/2111-08.
- [11] J.-D. Schmöcker, A. Fonzone, H. Shimamoto, F. Kurauchi, and M. G. Bell, "Frequency-based transit assignment considering seat capacities," *Transp. Res. B, Methodol.*, vol. 45, no. 2, pp. 392–408, Feb. 2011.
- [12] W. Szeto, Y. Jiang, K. Wong, and M. Solayappan, "Reliability-based stochastic transit assignment with capacity constraints: Formulation and solution method," *Transp. Res. C, Emerg. Technol.*, vol. 35, pp. 286–304, Oct. 2013.
- [13] W. Y. Szeto, M. Solayappan, and Y. Jiang, "Reliability-based transit assignment for congested stochastic transit networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 26, no. 4, pp. 311–326, May 2011.
- [14] C. A. O'Flaherty and D. O. Mangan, "Bus passenger waiting times in central areas," *Traffic Eng. Control*, vol. 11, pp. 419–421, 1970.
- [15] L. A. Bowman and M. A. Turnquist, "Service frequency, schedule reliability and passenger wait times at transit stops," *Transp. Res. A, Gen.*, vol. 15, no. 6, pp. 465–471, Nov. 1981.
- [16] W. Fan and R. B. Machemehl, "Do transit users just wait for buses or wait with strategies?" *Transp. Res. Rec.*, vol. 2111, no. 1, pp. 169–176, Jan. 2009, doi: 10.3141/2111-19.
- [17] W. Fan and R. B. Machemehl, "Characterizing bus transit passenger waiting times," *Mater. Spec. Conf. Can. Soc. Civ. Eng.*, vol. 7, pp. 1–10, Jun. 2002.
- [18] M. F. Nygaard and T. Tørset, "Waiting time strategy for public transport passengers," *Annu. Transp. Conf. Aalborg Univ., Mumbai, India, Tech. Rep.*, 2016, pp. 1–14.
- [19] M. Bagchi and P. R. White, "What role for smart-card data from bus system?" in *Proc. Inst. Civil Eng.-Municipal Eng.*, London, U.K., vol. 157, 2004, pp. 39–46.
- [20] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 292–306, May 2018.
- [21] L. Sun, D.-H. Lee, A. Erath, and X. Huang, "Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system," in *Proc. ACM SIGKDD Int. Workshop Urban Comput.*, Aug. 2012, pp. 142–148.
- [22] A. Tirachini, L. Sun, A. Erath, and A. Chakirov, "Valuation of sitting and standing in metro trains using revealed preferences," *Transp. Policy*, vol. 47, pp. 94–104, Apr. 2016.
- [23] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data," *Transp. Res. Rec.*, vol. 2275, no. 1, pp. 58–67, Jan. 2012.
- [24] X. Chen, L. Zhou, Y. Yue, Y. Zhou, and L. Liu, "Data-driven method to estimate the maximum likelihood space-time trajectory in an urban rail transit system," *Sustainability*, vol. 10, no. 6, p. 1752, May 2018.
- [25] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Syst. Appl.*, vol. 95, pp. 324–332, Apr. 2018.
- [26] L. Sun, Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen, "An integrated Bayesian approach for passenger flow assignment in metro networks," *Transp. Res. C, Emerg. Technol.*, vol. 52, pp. 116–131, Mar. 2015.
- [27] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "A probabilistic Passenger-to-Train Assignment Model based on automated data," *Transp. Res. B Methodol.*, vol. 104, pp. 522–542, 2017.
- [28] Y. Zhu, "Passenger-to-train assignment model based on automated data," M.S. thesis, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2014. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/90075>
- [29] Q. Zou, X. Yao, P. Zhao, F. Dou, and T. Yang, "Managing recurrent congestion of subway network in peak hours with station inflow control," *J. Adv. Transp.*, vol. 2018, Jan. 2018, Art. no. 6931025.
- [30] M. Ottomanelli, G. Iannucci, and D. Sassanelli, "Simplified model for pedestrian-vehicle interactions at road crossings based on discrete events system," *Transp. Res. Rec.*, vol. 2316, no. 1, pp. 58–68, Jan. 2012.
- [31] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. INFOCOM*, vol. 6, 2006, pp. 1–13.
- [32] R. Zhang, X. Zhu, and L. Lei, "Study on intersections with median BRT stations: Focused on pedestrian street-crossing characteristics," in *Proc. ICTP*, Jul. 2009, pp. 1392–1399.
- [33] W. Daamen and S. P. Hoogendoorn, "Free speed distributions for pedestrian traffic," in *Proc. TRB-Annu. Meeting*, Washington, DC, USA, 2006.
- [34] Y. Asakura, T. Iryo, Y. Nakajima, and T. Kusakabe, "Estimation of behavioural change of railway passengers using smart card data," *Public Transp.*, vol. 4, no. 1, pp. 1–16, Jul. 2012.
- [35] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers-train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 5, pp. 731–749, Sep. 2010.
- [36] L. Sun, J. G. Jin, D.-H. Lee, K. W. Axhausen, and A. Erath, "Demand-driven timetable design for metro services," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 284–299, Sep. 2014.



WANJUN LI received the B.S. and M.S. degrees in civil engineering from Iowa State University, Ames, IA, USA. She is currently pursuing the Ph.D. degree with the School of Traffic and Transportation, Beijing Jiaotong University, China. Her research interests include transportation big-data analysis, travel behavior analysis, metro planning, and traffic engineering.



XIAOMENG LI received the Ph.D. degree from Beijing Jiaotong University, in January 2017. She is currently a Research Fellow with the Centre for Accident Research and Road Safety–Queensland (CARRS-Q), Queensland University of Technology. Her researches mainly focus on road traffic safety, driving performance, big-data analysis, ITS, and traffic engineering.



XUEDONG YAN received the M.S. and Ph.D. degrees from the University of Central Florida, Orlando, FL, USA. He was the Director of traffic safety with the Transportation Research Center, Tennessee University, Knoxville, TN, USA. His research interests include road safety analysis, travel behavior and performance analysis, and transportation big-data analysis.

Dr. Yan is the Chief Editor of *Journal of Transportation Safety and Security*.



JINGSI YANG received the B.S. and M.S. degrees from Beijing Jiaotong University, China, where she is currently pursuing the Ph.D. degree with the School of Traffic and Transportation. Her major fields of study include traffic safety analysis, driving simulator analysis, and transportation big data analysis.

...