Hindawi
Complexity
Volume 2022, Article ID 5451017, 13 pages
https://doi.org/10.1155/2022/5451017

*Research Article*

# Data-Driven Method for Passenger Path Choice Inference in Congested Subway Network

**Guanghui Su**,[1] **Bingfeng Si**,[1] **Fang Zhao**,[1] **and He Li**[2]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[2]*Beijing Metro Network Control Center, Beijing 100101, China*

Correspondence should be addressed to Bingfeng Si; bfsi@bjtu.edu.cn

Received 30 September 2021; Revised 10 January 2022; Accepted 26 January 2022; Published 28 February 2022

Academic Editor: Harish Garg

Copyright © 2022 Guanghui Su et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a congested large-scale subway network, the distribution of passenger flow in space-time dimension is very complex. Accurate estimation of passenger path choice is very important to understand the passenger flow distribution and even improve the operation service level. The availability of automated fare collection (AFC) data, timetable, and network topology data opens up a new opportunity to study this topic based on multisource data. A probability model is proposed in this study to calculate the individual passenger's path choice with multisource data, in which the impact of the network time-varying state (e.g., path travel time) on passenger path choice is fully considered. First, according to the number and characteristics of OD (origin-destination) candidate paths, the AFC data among special kinds of OD are selected to estimate the distribution of passengers' walking time and waiting time of each platform. Then, based on the composition of path travel time, its real-time probability distribution is formulated with the distribution of walking time, waiting time, and in-vehicle time as parameters. Finally, a membership function is introduced to evaluate the dependence between passenger's travel time and the real-time travel time distribution of each candidate path and take the path with the largest membership degree as passenger's choice. Finally, a case study with Beijing Subway data is applied to verify the effectiveness of the model presented in this study. We have compared and analysed the path calculation results in which the time-varying characteristics of network state are considered or not. The results indicate that a passenger's path choice behavior is affected by the network time-varying state, and our model can quantify the time-varying state and its impact on passenger path choice.

## 1. Introduction

To alleviate the pressure on urban public transport caused by the increasing demand for urban travel, more and more cities have built large-scale subway networks, such as Shanghai, Beijing, Paris, Tokyo, especially in China, about 20 cities with subway operation mileage of more than 100 kilometers. Compared with other sustainable means of transportation [1–4], the subway has the characteristics of large capacity [5] and high reliability and is deeply welcomed by urban residents [6, 7]. While the subway brings convenient travel services to passengers [8], the expansion of the network and the influx of passenger flow have also brought new problems to the subway operation, such as crowding [9, 10], train utilization efficiency [11], ticket revenue allocation among operators [12, 13], and the uncertainty of passenger flow distribution caused by the diversity of passenger path choice. To improve service quality and operation efficiency, operators urgently need to know the distribution of passenger flow in time and space dimensions [14]. In a large-scale subway network, the estimation of passenger path is the basis of calculating passenger flow distribution, which aims to identify which line they actually take and at which stations they transfer. Therefore, estimating passenger path is a subject of great practical significance.

It is very difficult to estimate the passenger travel path in large-scale subway networks. On the one hand, in order to improve passengers' travel experience, the subway operation service usually adopts "seamless transfer mode"; that is, passengers do not need to tap-out/tap-in when transferring

between different lines. In this case, if there are multiple candidate paths between OD (origin-destination), the passenger's path choice cannot be observed [15] or recorded. On the other hand, passenger path choice is affected by many factors, including travel time, number of transfers, departure headway, walking time, waiting time, crowding. In order to depict the impact of these factors on passenger path choice, based on the utility maximization criterion, using questionnaires to calibrate the logit parameters is a widely used traditional method. Raveau et al. [16] proposed a multinomial logit model to study the path choice behavior of subway passengers, in which the utility function considers many factors that affect passengers' path choice. Later, in order to characterize the impact of the correlation between candidate paths on passenger's path choice, Raveau et al. [17] extended their research to a C-logit model including a "commonality factor," which can better deal with the calculation error caused by a path over-lapping. Literatures [12, 18] improved the calculation accuracy of the logit model by adding the transfer cost to the model and establishing multiclass utility function according to passenger category, respectively. However, there are many limitations in using a questionnaire to calibrate parameters, such as (1) collecting data by conducting surveys is costly both in time and resources [19], (2) due to the influence of survey location and respondents, the questionnaire results are often limited in scale and diversity [13], and (3) some key factors (such as congestion, comfortable) are difficult to quantify [20]. This limits the practical application of the logit model.

To overcome the limitations of traditional methods relying on questionnaires, scholars try to calibrate the parameters of the logit model by the data-driven method. Based on AFC (automated fare collection) data, Sun et al. [13] proposed an integrated Bayesian inference approach to study passenger path choice behavior. The core of this approach is still the logit model, while different from the traditional method, and it requires very limited information as input (e.g., passenger travel time from AFC data) but provides comprehensive posterior knowledge of passenger path choice. Then, Xu et al. [21] extended this model and considered the train crowding. However, the above studies did not consider the impact of time-dependent parameters (e.g., train timetable) and network time-varying state (e.g., path travel time) on passenger path choice.

In order to further consider the impact of time-dependence and network time-varying state on passenger path choice behavior, Zhou and Xu [22] combined AFC data and train timetable data to infer passengers' path choice. The gap between the passengers arriving at the platform at the fastest walking speed and the first train's departure that they can catch was defined as the "surplus time." Then, a probability function was constructed with the "surplus time" of each path's "boarding plan." Finally, the path with the highest probability was assigned to the passenger. However, the author does not consider the impact of waiting time distribution on path travel time, which is an important factor affecting passenger path choice. Similarly, Zhao et al. [23] proposed a probability model to convert the likelihood of passengers choosing different paths to the probability of taking different trains, and the method was verified with the data collected from the Shenzhen metro system. Li et al. [24] proposed a synchronous clustering method based on passenger "pure" travel time to calculate passenger path selection. The "pure" travel time refers to the remaining time after their access/egress walk time and waiting time are deleted from their travel time. However, deleting the waiting time may affect the calculation accuracy of this method in crowded subway network. Wu et al. [25] presented a density peaks clustering algorithm (DPCA) to infer passengers' path choice. First, passengers among the same OD are clustered by DPCA according to their travel time and then introduced a method to estimate the theoretical travel time of each candidate path considering the uncertain walking time and transfer time, and finally, according to the similarity between the cluster center value and the theoretical travel time of the candidate path, the cluster (with its passenger) and the candidate path are matched.

The above studies have made a great contribution to estimating passenger path choice based on data-driven method, but it can still be further improved. First, the probability distribution characteristics of time data are very important [26, 27], and these studies tend to rely on assumptions rather than statistical analysis. For example, it is assumed that passenger travel time follows normal distribution or uniform distribution, which will affect the accuracy and applicability of the model. Second, affected by passenger flow and crowding [10, 28] factors, network state (e.g., path travel time) has significant time-varying characteristics, which will ultimately affect the travel choice and travel time of passengers. Existing studies rarely explicitly consider the impact of this feature on passenger path choice. For example, there are two candidate paths for an OD pair, namely p1 and p2. The travel time of p1 in off-peak hours and peak hours is 5 minutes and 10 minutes, respectively, and the corresponding travel time of p2 is 10 minutes and 15 minutes, respectively. In this case, if only the passenger's travel time (e.g., 10 minutes) is used, it is difficult to accurately infer the passenger's path choice. If supplemented by passenger departure time (e.g., peak hours), then p1 is the most likely choice for passengers. It can be seen that the time-varying characteristics are helpful to estimate the passenger path.

To address the research gap mentioned above, this study proposes a probability model which can explicitly consider the influence of time-varying characteristics on estimating the passenger path. The model takes AFC, train timetable, and network topology data as inputs. First, the distribution characteristics of path travel time components such as walk time and waiting time are studied. Then, in order to characterize the time-varying state, a method to calculate the distribution of path real-time travel time is presented. Finally, a membership function that can consider both the passenger travel time and the real-time travel time distribution of the candidate path is introduced to infer the passenger path choice. To highlight the novelty of our work, some relevant studies are summarized in Table 1 and compared with our study, in terms of passenger travel time (A), the real-time travel time of path (B), the probability

Table 1: Characteristics of relevant works in comparison with our work.

| Publications | A | B | C | Model | Data |
|---|---|---|---|---|---|
| [13] | Yes | No | Assumption-based | Logit | AFC |
| [22] | Yes | No | Assumption-based | Probability model | AFC, timetable |
| [24] | Yes | No | Assumption-based | Clustering | AFC, timetable |
| [25] | Yes | No | Assumption-based | Clustering | AFC, timetable |
| [29] | Yes | No | Assumption-based | Logit | RP data + AFC |
| Our study | Yes | Yes | Data fitting, KS test | Probability model | AFC, timetable, network topology |

distribution of travel time (C), model and data. The detailed contributions of this study are summarized as follows:

(1) We proposed a multisource data-driven method to estimate passenger walking time and waiting time and tested the probability distribution characteristics of relevant time samples by the improved Kolmogorov-Smirnov (KS) method.

(2) According to the composition of path travel time and the distribution characteristics of walking time and waiting time, we presented a calculation method of path travel time distribution. On this basis, we further presented the calculation method of path real-time travel time distribution by calculating the time slot of waiting time distribution, which explicitly considers the time-varying state of the network.

(3) A membership function is proposed to evaluate the correlation 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 assigned to the passenger. In this way, the previous one-dimensional method that only considers passenger travel time can be extended to two dimensions; that is, both passenger travel time and passenger departure time are considered.

(4) Based on the data collected from Beijing Subway, the influence of time-varying characteristics of the route on estimating passenger path choice is analysed.

The rest of the paper is organized as follows: Section 2 states the problem and the necessary assumptions. Section 3 introduces the estimation method of walking time/waiting time probability distribution, first. On this basis, the general formula of path travel time distribution and passenger path estimation method is given. As an illustration, we apply the proposed model on the Beijing Subway network as a case study in Section 4 and compared the calculation results in which the time-varying characteristics of network states are considered or not. Finally, we conclude our study, summarize our main findings, and discuss future research directions in Section 5.

## 2. Problem Description

In the subway system, passengers' transaction information is recorded in the AFC system, specifically including the check-in time, check-out time, origin station, and destination station, but does exclude their transfer station and path information. Therefore, when there are multiple candidate paths between an OD pair, it is impossible to directly identify passenger path through AFC data. As shown in Figure 1, there are two paths from XD to CWM, and they are (1) XD-XWM-CWM and (2) XD-DD-CWM. They all include one transfer, and the travel distance is similar, so it is impossible to identify passengers' path choice directly.

The passenger travel time (obtained from the check-in and check-out time in AFC data) refers to the time spent by a passenger on a specific path between these OD. Hence, passenger travel time can be regarded as an observation of the travel time of this path. From a statistical point of view, when a random variable obeys a certain probability distribution, this variable should be within the confidence interval of the distribution. In other words, if we can know the travel time probability distribution of each candidate path, we can calculate the most likely path for passengers according to the confidence level of passenger travel time in the travel time distribution of each path.

The observation set is the basis for estimating the probability distribution of path travel time. For the ODs with only one candidate path, the travel time of passengers between this OD can be regarded as the travel time observation value of this unique path. However, when there are multiple paths between ODs, due to the lack of path information in AFC data, the travel time observation set of each candidate path between such ODs cannot be directly obtained. In other words, AFC data cannot directly estimate the path travel time distribution between such ODs. Therefore, how to estimate the path travel time distribution between multipath ODs through the single path OD (with a determined observation set) is the key to our study.

To that end, we first classify the OD in the network. According to the number of candidate paths and the path characteristics, the OD can be divided into three categories: (1) type-I OD, which means that there is only one candidate path, and passengers do not need to transfer in travel; (2) type-II OD refers to that there is only one candidate path but the O station and D station are not on the same line so that passengers have to transfer in travel; (3) type-III OD pair has multiple candidate paths. Take part of the Beijing Subway network in Figure 1 as an example; NLSL-FXM belongs to type-I OD; NLSL-FCM belongs to type-II OD; NLSL-CWM and FCM-JGM belong to type-III OD.

According to the above analysis, we know that the path travel time observation set of type-I OD can be obtained directly, so our goal is to infer the passenger path choice between type-III ODs with type-I ODs. To bridge these two types of OD, we need to analyse the travel time composition

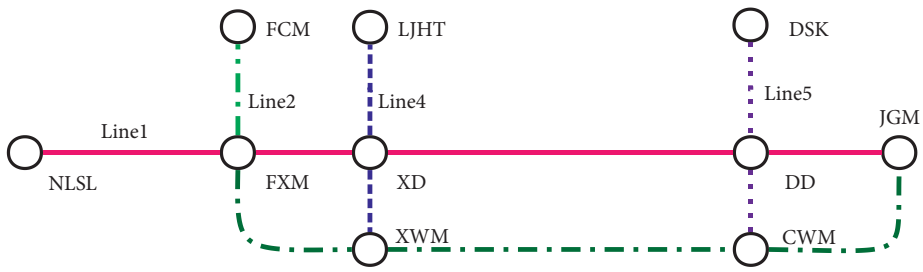


Figure 1: Partial diagram of Beijing Subway network.

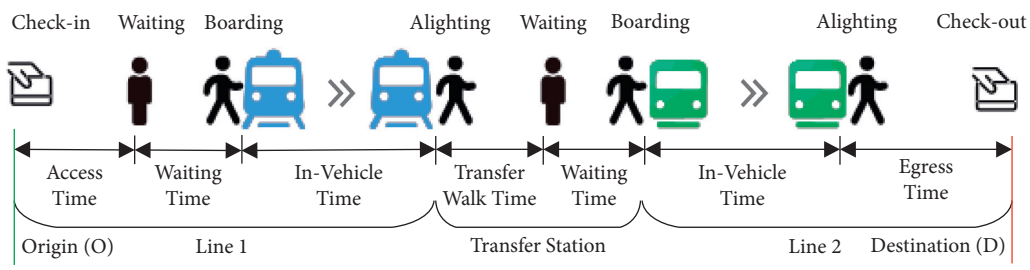


Figure 2: Passenger travel diagram by subway.

of passengers. As shown in Figure 2, passengers' travel time on a specific path includes in-vehicle time, walking time, and waiting time. That is, the path travel time of all kinds of OD is composed of these parts. Therefore, we can use the observation set between type-I OD to estimate the time components and then use these time components to estimate the travel time distribution of the path between type-III OD. Finally, the passenger path choice is estimated by means of the passenger travel time and the probability distribution of path travel time. Usually, the train runs according to the timetable, so the in-vehicle time is fixed; walking time can be divided into access time, transfer time, and egress time. The egress time can be calculated according to the timetable data and passenger check-out data. Nevertheless, it is unable to distinguish between waiting time and access/transfer time. Considering that the walking time is a linear function of walking distance and walking speed, if we assume that the walking speed of the same passenger in a trip is consistent, we can calculate the access time and transfer time according to the egress time and channel distances. This assumption is reasonable, although, in practice, some passengers may prolong their walking time due to channel congestion when check-in or transfer; the extended time can be regarded as part of the waiting time. The waiting time of passengers is affected by factors such as train departure frequency and the number of people waiting at the platform, which has strong randomness; that is, the waiting time at the platform has a significant time-varying characteristic. That is, each time component can be decomposed and estimated.

To sum up, based on AFC data, train timetable, and walking distance of subway platform, a passenger path estimation model considering time-varying travel time is proposed in this paper. First, based on type-I OD, the passenger walking time distribution and the waiting time distribution of each platform are estimated. Then, according to the path constituent elements, a method to restore the travel time distribution of each candidate path between type-III OD is proposed. Finally, taking the passenger check-in time, travel time and the real-time travel time distribution of each candidate path as parameters, a membership is introduced to estimate passenger path choice. The calculation flow is shown in Figure 3. To facilitate the subsequent analysis and modeling, the following assumptions are proposed:

(i) A1: Egress time is not affected by passenger traffic congestion but is only related to passenger walking speed and walking distance.

(ii) A2: The walking speed of the same passenger is consistent in different stages of a trip.

(iii) A3: The distance between the ticket gates and the platform is the same at the same station.

(iv) A4: Ignoring the time spent by passengers boarding and alighting the train.

## 3. Methodology

*3.1. Probability Distribution of Walking Time.* According to the previous analysis, walking time includes three kinds: access time, transfer time, and egress time. The egress time can be obtained from the passenger travel records between type-I OD. Therefore, the estimation method of egress time distribution is given first. Then, the calculation method of access time and transfer time distribution is derived based on the egress time distribution.

The passenger $i$ between type-I OD pair $u-v$, whose check-in time at platform $u$ and check-out time from the platform $v$ are $t_i^{\text{tapin}}$ and $t_i^{\text{tapout}}$, respectively, and Figure 4 shows all the possible trips of this passenger. The horizontal axis is the passenger's travel time, and the vertical axis represents their location. The access time is the time passenger spends from the access ticket gate to the origin

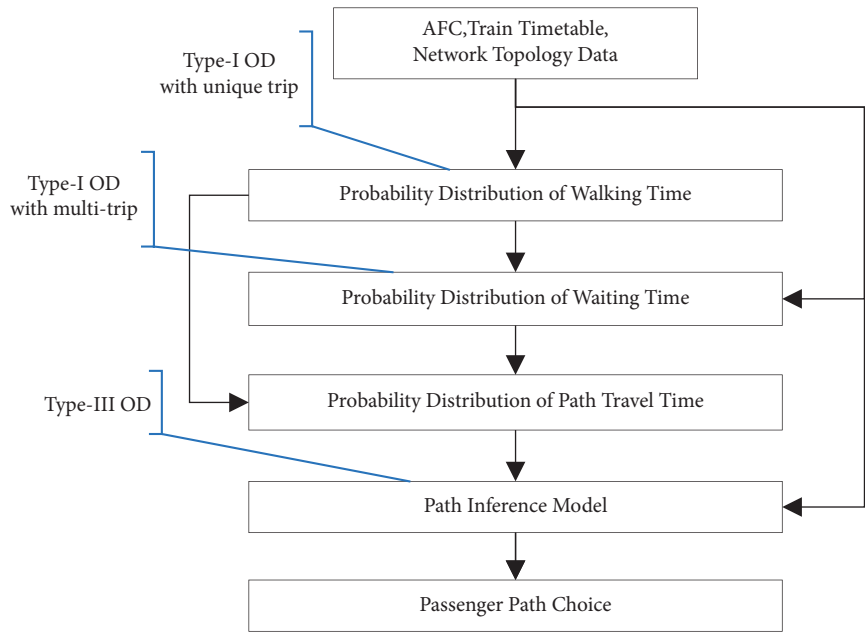


Figure 3: Calculation flow of passenger path choice.

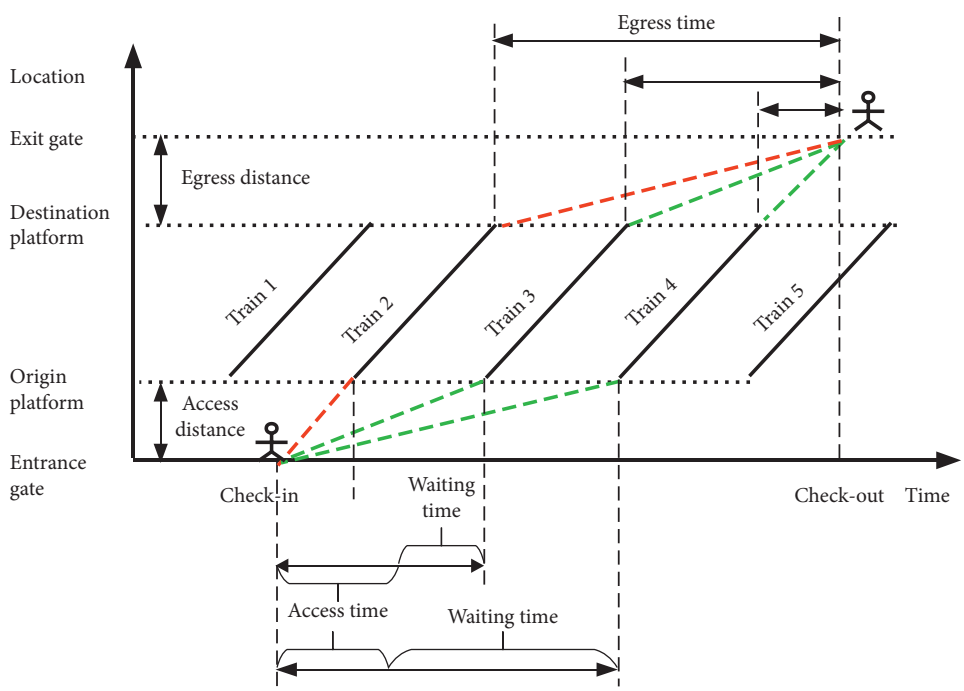


Figure 4: Diagram of passenger trips between Type-I OD

platform. Waiting time is the dwell time passenger spends at the platform before boarding. Denote $t_{j,v}^{\text{arr}}$ as the arrival time of train $j \in J$ to the platform $v$, and the egress time is equal to the time spent between $t_{j,v}^{\text{arr}}$ and $t_i^{\text{tapout}}$. It can be seen from Figure 4 that there are three potential trips (2,3,4) run in the range of $t_i^{\text{tapin}}$ and $t_i^{\text{tapout}}$, but the access/egress time reserved for passenger $i$ by train 2 does not meet the walking speed consistency assumption (see assumption "A2"), so only train 3 and 4 are feasible.

When passenger $i$ has more than one feasible trip, each trip $j$ corresponds to an egress time $\tau_{i,j,v}^{\text{e}}$, namely:

$$\tau_{i,j,v}^{e} = t_i^{\text{tapout}} - t_{j,v}^{\text{arr}}, \quad \tau_{i,j,v}^{e} > 0. \tag{1}$$

According to the previous assumption that the walking speed of passengers is consistent when egress time is equal to $\tau_{i,j,v}^{\text{e}}$, the corresponding access time $\tau_{i,j,u}^{\text{a}}$ can be expressed as follows:

$$\tau^{a}_{i,j,u} = \tau^{e}_{i,j,v} \cdot \eta_{u,v}, \quad \eta_{u,v} > 0, \tag{2}$$

where $\eta_{u,v} = d^{\mathrm{a}}_{u}/d^{\mathrm{e}}_{v}$ is the distance ratio to the entrance channel and the exit channel, and $d^{\mathrm{a}}_{u}$ and $d^{\mathrm{e}}_{v}$ are the entrance channel distance of the platform $u$ and the exit channel distance of platform $v$, respectively.

Obviously, if the trip $j$ is feasible for passenger $i$, its departure time $t^{\mathrm{dep}}_{j,u}$ from platform $u$ must satisfy

$$t^{\mathrm{tapin}}_{i} + \tau^{\mathrm{a}}_{i,j,u} \le t^{\mathrm{dep}}_{j,u}. \tag{3}$$

Meanwhile, from Figure 4, it can be seen that if passenger $i$ has only one feasible trip, the egress time can be uniquely determined by the equation (1). Therefore, based on such passengers' egress time, the egress time probability distribution of the corresponding platform can be fitted. However, having only one feasible trip means that passengers' check-out time should be between the arrival time of his/her unique feasible trip $j$ and of neighbour follow-up train $j+1$, namely:

$$t^{\mathrm{arr}}_{j,v} < t^{\mathrm{tapout}}_{i} \le t^{\mathrm{arr}}_{j+1,v}. \tag{4}$$

Equation (4) means that the upper limit of such passengers' egress walking time cannot exceed the trains' headway when passengers leave the departure platform, but this may lead to a deviation in the estimation of walking time. In Beijing Subway, the off-peak hours' headway is about 4 minutes, which is long enough for most passengers to walk from the platform to the ticket gate. Therefore, using the AFC data with only one feasible trip in the off-peak hours can effectively reduce the deviation of estimating the passenger's egress time. Since the walking speed changes relatively small in different periods, we take the off-peak hours' walking time distribution to represent the whole day's walking time distribution.

Assuming that all passengers between type-I OD $u-v$ satisfying equations (1) and (3), and having one feasible trip during off-peak hours, the sample space of their egress walking time is $\Gamma^{e}_{v} = \left\{\tau^{e}_{i,j,v} | i \in I, j \in J \& j+1 \notin J\right\}$. Referring to the existing research [13, 27, 30, 31], this study uses the normal distribution [13, 30], logarithmic normal distribution [31], and gamma distribution [27] to fit the egress time of the 644 platforms in the Beijing Subway network and uses the improved KS (Kolmogorov-Smirnov) method [32] to test the fitting result. Table 2 gives the corresponding KS test statistical results.

It can be seen that the lognormal distribution has the best fitting effect on the sample space. Therefore, it is safe to assume that passengers' egress time follows lognormal distribution $\Gamma^{e}_{v} \sim \mathrm{Log}N(\mu^{e}_{v}, \sigma^{e}_{v})$, where $\mu^{\mathrm{e}}_{v}$ and $\sigma^{\mathrm{e}}_{v}$ are the logarithmic mean and standard deviation of the egress walking time, which can be estimated from the egress time sample space by using the MLE (maximum likelihood estimation) method. In this way, we can get the occurrence probability $P(\tau^{\mathrm{e}}_{i,j,v})$ of the egress time $\tau^{\mathrm{e}}_{i,j,v}$ by discretizing the fitted probability distribution density; that is,

$$P\left(\tau^{e}_{i,j,v} - \varepsilon < t \le \tau^{e}_{i,j,v} + \varepsilon\right) = \int_{\tau^{e}_{i,j,v}-\varepsilon}^{\tau^{e}_{i,j,v}+\varepsilon} f(t)\mathrm{d}t, \tag{5}$$

where $\varepsilon$ is the time granularity, such as 1 second; $f(t)$ is the probability density function of the egress time, which is in the following form:

$$f(t) = \frac{1}{t\sqrt{2\pi}\sigma^{\mathrm{e}}_{v}} \exp\left(-\frac{\left(\ln t - \mu^{\mathrm{e}}_{v}\right)^{2}}{2\left(\sigma^{\mathrm{e}}_{v}\right)^{2}}\right). \tag{6}$$

Suppose the influence of channel congestion and other factors on the access time is not considered. In that case, according to equation (2), the access time sample space $\Gamma^{a}_{u}$ of the platform can be expressed as the scalar multiplication of the walking distance coefficient and the sample space $\Gamma^{e}_{v}$ of the egress time.

Suppose $\Gamma^{a}_{u}$ obeys lognormal distribution $\Gamma^{a}_{u} \sim \mathrm{Log}N(\mu^{a}_{u}, \sigma^{a}_{u})$, where $\mu^{a}_{u}$ and $\sigma^{a}_{u}$ are the logarithmic mean and standard deviation of the access time of platform $u$. In practice, the extension of access/transfer walking time caused by congestion can be regarded as a part of the subsequent waiting time connected with it. Therefore, this treatment will not affect the estimation of feasible trips.

As a special kind of access time (the walking time passengers spend between the arrival platform and the next departure platform), the main factor affecting the transfer time is also the channel distance. Therefore, the distribution $\Gamma^{c}_{u1,u2} \sim \mathrm{Log}N\ (\mu^{c}_{u1,u2}, \sigma^{c}_{u1,u2})$ of transfer time from the platform $u1$ to the platform $u2$ can be calculated by the same method, where $\mu^{c}_{u1,u2}$ and $\sigma^{c}_{u1,u2}$ are the logarithmic mean and standard deviation of the walking time from the platform $u1$ to the platform $u2$.

*3.2. Probability Distribution of Waiting Time.* The uncertainty of platform waiting time is mainly manifested in two aspects. On the one hand, the waiting time at the same platform varies in different periods. On the other hand, different platforms' waiting time is different in the same period. The waiting passengers include the new check-in passengers, transfer passengers, and stranded passengers. Generally speaking, the boarding probability of the waiting passengers on the platform has nothing to do with their types. The earlier the passengers arrive at the platform, the easier for them to choose the vantage position and more likely to board, especially when the arriving train's remaining capacity is not enough to load all of them. Therefore, the waiting time distribution of new passengers can represent all passengers' waiting time distribution on the platform.

It can be seen from Figure 4 that when passenger $i$ takes the trip $j$, his/her waiting time $\tau^{\mathrm{w}}_{i,j,u}$ at the departure platform $u$ is

$$\tau^{w}_{i,j,u} = t^{\mathrm{dep}}_{j,u} - \left(t^{\mathrm{tapin}}_{i} + \tau^{a}_{i,j,u}\right), \quad \tau^{w}_{i,j,u} \ge 0,\ j \in J_{i}. \tag{7}$$

Substituting equations (1) and (2) into the above equation, we can get the following results:

Table 2: The KS test statistical result of egress time distribution.

| | Normal distribution | Lognormal distribution | Gamma distribution |
|---|---|---|---|
| Count | 364 | 450 | 432 |
| Ratio | 56.52% | 69.88% | 67.08% |

$$\tau^{\mathrm{w}}_{i,j,u} = t_i^{\mathrm{tapout}} - t_i^{\mathrm{tapin}} - \tau^{\mathrm{e}}_{i,j,v}\left(1 + \eta_{u,v}\right) - \tau^{\mathrm{b}}_{j,u,v}, j \in J_i, \tag{8}$$

where $\tau^{\mathrm{b}}_{j,u,v} = t^{\mathrm{arr}}_{j,v} - t^{\mathrm{dep}}_{j,u}$ is the running time of train $j$ from platform $u$ to platform $v$. In an urban rail transit network, the train's running time between two platforms on the same line is usually fixed, that is, $\tau^{b}_{j,u,v} = \tau^{b}_{u,v} = C$. Therefore, in equation (8), $\tau^{\mathrm{w}}_{i,j,u}$ can be abbreviated as the function of $\tau^{\mathrm{e}}_{i,j,v}$.

Let $P(\tau^{w}_{i,j,u})$ be the waiting time probability of passenger $i$ at the platform $u$ when he/she takes trip $j$. From the above analysis, we can know that $P(\tau^{w}_{i,j,u})$ equals to the likelihood of passengers' corresponding egress walking time; that is,

$$P\left(\tau^{\mathrm{w}}_{i,j,u}\right) = P\left(\tau^{\mathrm{e}}_{i,j,v} = t_i^{\mathrm{tapout}} - t^{\mathrm{arr}}_{j,v}\right), j \in J_i, \tag{9}$$

where the egress walking time probability $P(\tau^{\mathrm{e}}_{i,j,v})$ is given by equation (5).

Equation (9) is normalized, that is, $P'(\tau^{\mathrm{w}}_{i,j,u}) = P(\tau^{\mathrm{w}}_{i,j,u})(\sum_{j=1}^{J_i} P(\tau^{\mathrm{w}}_{i,j,u}))^{-1}$.

The discrete value of waiting time is equal to the product of waiting time generated by different trips and its corresponding normalized probability. Therefore, the sample space of the platform waiting time can be expressed as follows:

$$\Gamma^{w}_{u} = \left\{\sum\nolimits_{j=1}^{J_i} P'\left(\tau^{w}_{i,j,u}\right)\tau^{w}_{i,j,u} | i \in I\right\}. \tag{10}$$

Since the platform waiting time changes rapidly, we split the operation time by a short time (such as 30 minutes), and then, we get 22600 periods in total for 644 platforms in the Beijing Subway network. Similarly, the waiting time distribution is fitted based on different probability distributions [13, 27, 30, 31], and the fitting results are tested by improved KS method [32]. Table 3 shows the KS test statistical results of the waiting time distribution.

We can know that lognormal distribution is the best fit for samples, so we assume that the waiting time at platform $u$ follows the lognormal distribution with $\mu^{w}_{u}$ and $\sigma^{w}_{u}$ as logarithmic mean and standard deviation, and the two parameters can be estimated from the sample space $\Gamma^{w}_{u}$. For the platform with fewer passengers, a fixed value (such as the corresponding period's headway) is used as the waiting time for passengers at the platform.

*3.3. Probability Distribution of Path Travel Time.* The above analysis obtains the travel time distribution of each part of the path. For the convenience of expression, we number each section/subpath's travel time in chronological order. So, if there are $k \in K$ candidate paths between any type-III OD, according to Figure 2, the travel time of path $k$ can be written as the sum of each subpaths' travel time.

$$\tau_k = \sum_{l=1}^{L_k} \tau_{k,l}, \tag{11}$$

where $l \in L_k$ is the sections' order number of path $k$.

According to the previous analysis, the subpath travel time $\tau_{k,l}$ can be regarded as independent of each other. Beaulieu et al. [33] pointed out that when random variables obey lognormal distribution, the sum of random variables also obeys lognormal distribution. Therefore, the travel time $\tau_k$ of path $k$ follows the lognormal distribution with $\mu_k$ and $\sigma_k$ as parameters, and $\mu_k$ and $\sigma_k$ are equal to the logarithmic mean and standard deviation of $\tau_k$, respectively.

According to the Wilkinson method [33], the first/second moments of the path travel time equal the first/second moments of the sum of the subpaths' travel time, respectively. Therefore,

$$E(\tau_k) = E\left(\sum_{l=1}^{L_k} \tau_{k,l}\right), \tag{12}$$

$$E\left((\tau_k)^2\right) = E\left(\left(\sum_{l=1}^{L_k} \tau_{k,l}\right)^2\right). \tag{13}$$

According to the nature of lognormal distribution, there are

$$E(\tau_k) = \exp\left(\mu_k + \frac{\sigma_k^2}{2}\right), \tag{14}$$

$$E\left((\tau_k)^2\right) = \exp\left(2\mu_k + 2\sigma_k^2\right). \tag{15}$$

Substituting equations (14) and (15) into equations (12) and (13), respectively.

$$\begin{aligned} \exp\left(\mu_k + \frac{\sigma_k^2}{2}\right) &= \sum_{l=1}^{L_k}\left(E\left(\tau_{k,l}\right)\right), \\ \exp\left(2\mu_k + 2\sigma_k^2\right) &= \sum_{l=1}^{L_k}\left(E\left(\tau_{k,l}\right)\right)^2 \\ &\quad + \operatorname{cov}\left(\sum_{l=1}^{L_k}\tau_{k,l}, \sum_{l=1}^{L_k}\tau_{k,l}\right). \end{aligned} \tag{16}$$

Since the subpaths' travel time is independent of each other, their covariance is 0. Therefore, we can derive $\mu_k$ and $\sigma_k$ as follows:

Table 3: The KS test statistical result of waiting time distribution in different periods.

| | Normal distribution | Lognormal distribution | Gamma distribution |
|---|---|---|---|
| Count | 11662 | 16334 | 16170 |
| Ratio | 51.60% | 72.27% | 71.55% |

$$\sigma_k^2 = \ln\left(\frac{\sum_{l=1}^{L_k}\left[\left(\exp\left(\sigma_{k,l}^2\right)-1\right)\exp\left(2\mu_{k,l}+\sigma_{k,l}^2\right)\right]}{\left(\sum_{l=1}^{L_k}\delta_l\exp\left(\mu_{k,l}+\left(\sigma_{k,l}^2/2\right)\right)\right)^2}+1\right),$$
$$\delta_l = \begin{cases} 1, & \text{if } \sigma_{k,l} \neq 0, \\ 0, & \text{if } \sigma_{k,l} = 0, \end{cases} \tag{17}$$

$$\mu_k = \ln\left(\sum_{l=1}^{L_k} xp\left(\mu_{k,l}+\frac{\sigma_{k,l}^2}{2}\right)\right)-\frac{\sigma_k^2}{2}, \tag{18}$$

where $\mu_{k,l}$ and $\sigma_{k,l}$ are the logarithmic expectation and standard deviation of subpaths' travel time.

The probability density function of path travel time can be expressed as

$$f\left(\tau_k,\mu_k,\sigma_k\right) = \frac{1}{\sigma_k\tau_k\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\ \tau_k-\mu_k\right)^2}{2\left(\sigma_k\right)^2}\right). \tag{19}$$

*3.4. Path Inference Method.* The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

Before estimating the passenger's path choice, it is necessary to estimate the travel time distribution of each candidate path. Suppose the theoretical travel time of path $k \in K$ between a type-III OD is $\tau_k$, and according to equation (11), $\tau_k$ is the sum of in-vehicle time, walking time, and waiting time under ideal conditions, so it is usually shorter than the actual travel time of path $k$. Hence, if path $k$ was the feasible path of passenger $i$, his/her travel time should be not less than $\tau_k$, namely:

$$\tau_k \leq t_i^{\text{tapout}} - t_i^{\text{tapin}}. \tag{20}$$

To calculate the real-time travel time distribution of the path $k$, it is necessary to estimate the real-time waiting time distribution of each departure platform, that is, to calculate the time slot when passengers arrive at different platforms. Assuming that path $k$ includes $m \in M$ departure platforms, passenger $i$ arrives at the $m$th departure platform at time $t_{k,m}^{\text{w}}$, and the time offset relative to $t_{k,m}^{\text{w}}$ is $\Delta\tau_{k,m} = t_{k,m}^{\text{w}} - t_i^{\text{tapin}}$ (see Figure 4). Denote $L_{k,m}$ as the number of subpaths before departure platform $m$; then from (11), the offset can be approximated as the sum of the mean of $L_{k,m}$ subpaths' travel time before he/she arriving at platform $m$, namely:

$$\Delta\tau_{k,m} \approx \sum_{l=1}^{L_{k,m}} E\left(\tau_{k,l}\right). \tag{21}$$

The interesting time period is divided into $S$ equal time slots, and the length of each one is $\Delta s$. Denote $s_m$ as the time period of passenger $i$ arrives at platform $m$, so $s_m = \left\lceil\left(t_i^{\text{tapin}}+\Delta\tau_{k,m}\right)\cdot\Delta s^{-1}\right\rceil$; therefore, the waiting time distribution of the platform $m$ follows $\tau_{m,s_m}^{\text{w}} \sim \text{Log}N\left(\mu_{m,s_m}^{\text{w}},\sigma_{m,s_m}^{\text{w}}\right)$, $\mu_{m,s_m}^{\text{w}}$ and $\sigma_{m,s_m}^{\text{w}}$ are the logarithmic mean and standard deviation of the waiting time of platform $m$ in the period $s_m$. Substituting them into equations (18), and (19), the real-time travel time distribution parameters $\mu_{k,t_i^{\text{tapin}}}$ and $\sigma_{k,t_i^{\text{tapin}}}$ of path $k$ at time $t_i^{\text{tapin}}$ can be obtained.

According to the above analysis, there is a specific dependence between passengers' travel time and the candidate paths' real-time travel time. And the membership function in fuzzy set theory can better deal with this situation. For example, literature [34] proposed a fuzzy power Heronian function to deal with multicriteria decision-making problems. Considering that the passenger travel time of the same OD pair obeys the lognormal distribution, this study constructs a lognormal distribution membership function in the following form to estimate the similarity between the passenger's travel time $\tau_i^{AFC}$ and the real-time travel time distribution of each candidate path. That is, the membership degree is equal to the discrete value of the passenger travel time on the probability distribution of each candidate path.

$$F_k\left(\tau_i^{\text{AFC}}\right) = P\left(\tau_i^{\text{AFC}}-\varepsilon < t \leq \tau_i^{\text{AFC}}+\varepsilon\right) = \int_{\tau_i^{\text{AFC}}-\varepsilon}^{\tau_i^{\text{AFC}}+\varepsilon} f_k(t)\mathrm{d}t,$$
$$f_k(t) = \frac{1}{t\sqrt{2\pi}\sigma_{k,t_i^{\text{tapin}}}}\exp\left(-\frac{\left(\ln\ t-\mu_{k,t_i^{\text{tapin}}}\right)^2}{2\left(\sigma_{k,t_i^{\text{tapin}}}\right)^2}\right), k \in K, \tag{22}$$

where $\varepsilon$ is the time granularity, such as 1 second; $\mu_{k,t_i^{\text{tapin}}}$ and $\sigma_{k,t_i^{\text{tapin}}}$ are the logarithmic mean and standard deviation of real-time travel time distribution of path $k$ at time $t_i^{\text{tapin}}$, respectively. Denote $\Delta_{i,k} = \left|\ln\ t-\mu_{k,t_i^{\text{tapin}}}\right|$, which represents the difference between the logarithm of passenger travel time and the logarithmic mean of the path real-time travel time, the smaller $\Delta_{i,k}$ is, the greater $f_k(t)$ is. When $\Delta_{i,k}$ is the same, the greater $\sigma_{k,t_i^{\text{tapin}}}$ is, the greater $f_k(t)$ is, indicating that the passenger travel time has higher confidence in the real-time travel time distribution of path $k$.

According to the principle of maximum membership degree, the passenger path between type-III OD pairs considering the time-varying characteristics of network state can be determined by the following formula:

$$k = \arg\max\left\{F_k\left(\tau_i^{AFC}\right)\right\}, k \in K. \tag{23}$$

## 4. Case Study and Comparison

For the purpose of model illustration and verification, we apply the proposed model to Beijing Subway network. The

network (as of 2017) consists of 19 lines with 608 km, serving 370 stations including 56 transfer stations. It serves about 5.4 million trips per day. Most of the passengers use smart card or mobile phone to pay the ticket, and the transactions would be recorded by the AFC systems, including the check-in and check-out stations and corresponding times.

We use a typical weekday AFC in October 2017 for the model application. As the punctuality rate of train operation exceeds 99.9% (refer to the report of Beijing Rail Transit Operation Co., Ltd.), we choose the planned timetable on the same day as AFC data as the input timetable. Walking distance is mainly obtained by two methods: (1) Baidu map (map.baidu.com) and (2) field investigation.

We select a typical OD pair (TTYB-CYM) for analysis and comparison, as shown in Figure 5. The relevant walking distances have been indicated in the figure ($A$ represents the access walking distance, $T$ represents the transfer distance, and $E$ represents the egress walking distance). Yen's algorithm [35] is adopted to generate candidate paths. To generate a high-quality path set in the complex Beijing Subway network, we introduce two auxiliary rules to Yen's algorithm, which are (1) dominance coefficient, that is, the ratio of travel time of path $k$ to travel time of path $k-1$ should be within a reasonable threshold, and otherwise, stop; (2) relative transfer number limit, that is, compared with the transfer times of the shortest circuit, the transfer number of path $k$ should be within a reasonable threshold, and otherwise, stop. In this case, we set the dominance coefficient and relative transfer number as 1.5 and 2, respectively. Hence, there are three candidate paths between this OD, namely:

Path 1: TTYB- > YHG (Line 5 to Line2)- > CYM; Transfer 1 time, the theoretical travel time is about 2179 seconds.

Path 2: TTYB- > DS (Line 5 to Line6)- > CYM; Transfer 1 time, the theoretical travel time is about 2430 seconds.

Path 3: TTYB- > LSQ (Line 5 to Line13)- > DZM (Line 13 to Line2)- > CYM; Transfer 2 times, the theoretical travel time is about 2540 seconds.

*4.1. Waiting Time Estimation.* The platform waiting time is an important indicator to measure the performance of the subway network. It can reflect the network congestion and passengers' dwell time in different time periods. Figure 6 shows the average waiting time at different transfer platforms between TTYB-CYM. The observation period is 6:00 ~ 16:00, and the peak hours are 7:00 ~ 9:00.

We found that in the peak hours (7:00 ~ 9:00), the waiting time of YHG and LSQ is close to the train headway, respectively. That is, most passengers can board the first train they can catch after they arrive at the platform. The waiting time of DS and DZM is significantly longer than their train headway, which indicates that passengers may be stranded. This may be caused by the platform overcrowding or the limited capacity of arriving trains. Whatever the reason, passengers' choice behavior will be affected. In addition, from the whole time period, the changing trend of waiting time and train headway is not consistent, indicating that the waiting time is affected by many factors and has an obvious random characteristic.

*4.2. Path Inference Results and Comparison.* Figure 7 illustrates the paths' travel time and the result of passenger path inference in the observation period. The horizontal axis represents the observation period, and the vertical axis represents the time. The solid line represents the travel time of each candidate path between TTYB-CYM, and the dot/triangle/diamond represents the passengers' path choice. For example, the dot indicated by the red arrow is the AFC data of a passenger and the path inference result. His/her check-in time and travel time are 7:05:31 and 2487 s, respectively, and the choice estimated by our model is path 1.

We can observe that (1) the travel time of path 1 and path 2 increases significantly during peak hours, while that of path 3 changes less. (2) Most passengers choose path 1 and path 2 with a short travel time, and few passengers choose path 3, which is consistent with the field survey investigation. (3) The distribution of passenger travel time in peak hours is compact, while that is relatively scattered in other periods. In the whole observation period, the average passenger travel time in peak hours is significantly longer than that in off-peak hours. Therefore, it is necessary to take the time-varying characteristics of path travel time into consideration when calculating passenger path choice.

Next, two comparative scenarios are designed to further analyse the impact of time-varying characteristics on estimating passenger path choice. In scenario 1, we apply our model to the Beijing Subway network. In scenario 2, we apply a static model to estimate passengers' path choice, in which the time-varying waiting time is replaced by the line's headway. Table 4 shows the comparison of passenger path inference results under two different scenarios.

Scenario 1: The calculation result shows that more than 90% of passengers tend to choose path 1 and path 2 with a shorter travel time and fewer transfer times during peak hours. And passengers prefer path 1 to path 2. Echoing this result, it can be seen from Figures 5 and 6 that the transfer distance and the waiting time of path 1 (platform YHG) are better than those of path 2 (platform DS).

Scenario 2: The result shows that more passengers choose path 2 during peak hours, which is inconsistent with the field survey. The reason for this phenomenon is that the waiting time in peak hours is much longer than the headway (Figure 6). Hence, if the time-varying waiting time is not considered, the passengers' actual travel time will be underestimated, which leads to the illusion that passengers prefer the path with a longer time. In the off-peak hours, the proportion of choosing path 1 increases but still less than 50%, while the proportion of passengers choosing path 3 is close to 20%, which is not reasonable either. According to our field survey, the transfer distance and travel time of path 3 are very unfriendly to passengers, so 20% is not a reliable proportion. This indicates that the waiting time is a key component to path travel time, which is very important for estimating the passenger path.

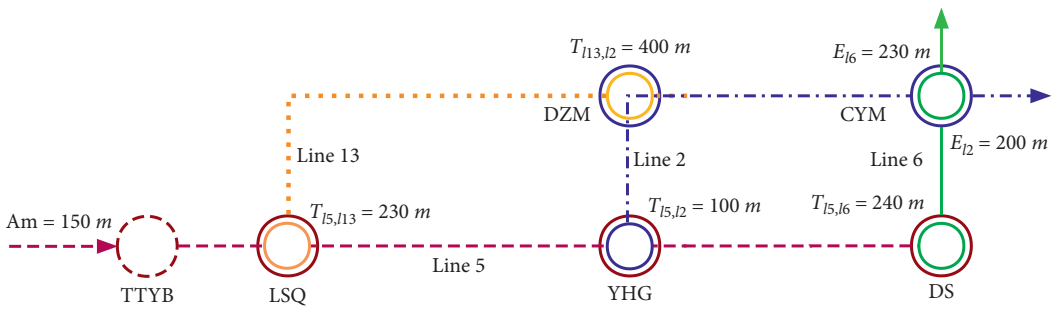


Figure 5: Candidate paths between TTYB and CYM

Figure 6: Average waiting time at transfer platforms.

In summary, no matter in peak hours or off-peak hours, the proportion of passengers choosing path 1 and path 2 travel exceeds 90% in both scenarios, which indicates that passengers tend to choose the path with fewer transfer times and shorter travel time. In scenario 1, the proportion of passengers who choose path 1 in peak hours is lower than that in off-peak hours, while the changing trend of passenger proportion of the other two paths is opposite. According to the user equilibrium, all passengers tend to choose the "advantage" path with shorter travel time and fewer transfer times, which leads to a sharp increase in waiting time and congestion/crowding of the "advantage" path. Then, the passengers who are sensitive to congestion/crowding or waiting time will seek the so-called "disadvantage" path instead, until no passengers can find a better path than their current one. In other words, in the peak hours, due to the waiting time and congestion/crowding, the attraction of line 1 is reduced, so some passengers turn to choose path 2 or path 3 instead. At this time, passengers' path choices reach an "equilibrium" state. In off-peak hours, with the decrease of passenger flow, the influence of waiting time and congestion/crowding on path 1 is weakened, so the passenger

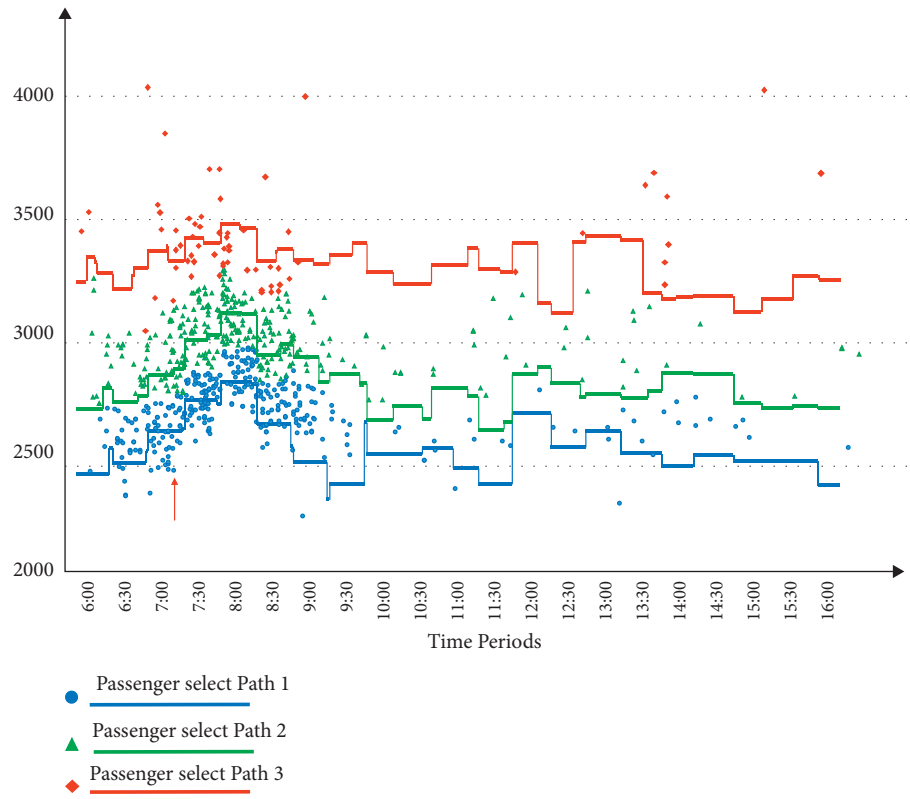


Figure 7: Path travel time and passenger path inference results.

Table 4: Comparison of path inference results.

| Methods | | Considering the time-varying state of the network | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes (scenario 1) | | | No (scenario 2) | | |
| Path | | Path 1 (%) | Path 2 (%) | Path 3 (%) | Path 1 (%) | Path 2 (%) | Path 3 (%) |
| Proportion | Peak hours | 53.31 | 38.75 | 7.94 | 30.62 | 66.16 | 3.21 |
| | Off-peak hours | 69.29 | 24.07 | 6.64 | 44.40 | 35.68 | 19.92 |

path choice reaches a new "equilibrium" state. That is, in off-peak hours, the proportion of choosing path 1 is higher than that in peak hours. In short, passenger path choice is a complex time-varying dynamic equilibrium. When calculating their path, the influence of time-varying network state on passenger path choice must be considered.

## 5. Conclusions

In this study, we propose a probability model to infer passenger path choice in which the time-vary characteristics of network state are explicitly considered. The model takes AFC, train timetable, and network topology data as input parameters and can provide a number of network performance indicators including passenger path choice, waiting time distribution, path travel time probability distribution. To that end, we propose a method to estimate the real-time travel time probability distribution of path based on multisource data. On this basis, a logarithmic membership function that can simultaneously consider passenger travel time and network state is introduced to infer the passenger path choice. Finally, a case study was conducted with the data collected from Beijing Subway to demonstrate the effectiveness of our proposed model. The case study shows that (1) the path travel time has significant time-varying characteristics (such as the change of waiting time); (2) compared with the method without considering the time-varying characteristics, the model proposed in this study can estimate the passenger path choice more accurately.

In general, this study fills the gaps in the existing studies from the following two aspects: (1) a method for estimating the probability distribution of walking time and waiting time based on multisource data is presented, and the probability distribution is tested by improved KS method [32]. The results show that the passenger travel time in Beijing Subway follows a lognormal distribution. (2) Time-varying state is explicitly considered in the estimation method of passenger path choice; that is, compared with the existing studies, the proposed model can consider both passenger travel time and the network state when passenger departs.

Practically, it is easy to apply the proposed model to the path choice or passenger flow calculation in the subway

system. Once the input information (e.g., AFC and other data) is predetermined, the subway network performance indicators such as path choice proportion and path travel time distribution can be obtained quickly. Specifically, it takes 12 minutes to estimate the path for the passengers of one day in Beijing Subway. Compared with traditional methods, on the one hand, this method is convenient and efficient, so it can meet the needs of daily operation; on the other hand, because this method is based on actual data rather than questionnaire, it can bring more operation benefits and managerial insights from the "current facts."

Nevertheless, this study still has some limitations: (1) in the case study, we use the train timetable rather than the real train movement data, which will still affect the accuracy of the method to a certain extent. (2) When the travel volume of the station is small (e.g., the suburban station at night), there may be a large error between the estimated value of waiting time distribution and the actual value. In this case, the estimated value can be replaced by the headway of the corresponding time period or corrected by the multiday cumulative AFC data of the station. (3) In this study, only one membership function is used to infer the passenger path choice. In the subsequent work, other advanced decision functions can be introduced for comparisons, such as fuzzy (monotone) measure [36] and Fermatean fuzzy group decision-making [37].

In the future, we will expand our research in the following aspects: (1) we study the data-fusion-based passenger path estimation model combined with questionnaire data or video data, and (2) we can also study the probability for passengers taking different trains on the basis of current work, so as to better estimate train utilization, platform crowding, etc. In addition, with the increasing scale of subway network, it is also an interesting direction to study passenger travel behavior based on network characteristics [38] or network topology index [39].

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The research was funded by the National Natural Science Foundation of China (91746201).

## References

[1] D. Pamucar, F. Ecer, and M. Deveci, "Assessment of alternative fuel vehicles for sustainable road transportation of United States using integrated fuzzy FUCOM and neutrosophic fuzzy MARCOS methodology," *The Science of the Total Environment*, vol. 788, Article ID 147763, 2021.

[2] R. Krishankumar, D. Pamucar, M. Deveci, and K. S. Ravichandran, "Prioritization of zero-carbon measures for sustainable urban mobility using integrated double hierarchy decision framework and EDAS approach," *The Science of the Total Environment*, vol. 797, Article ID 149068, 2021.

[3] D. Pamucar, M. Deveci, F. Canıtez, T. Paksoy, and V. Lukovac, "A novel methodology for prioritizing zero-carbon measures for sustainable transport," *Sustainable Production and Consumption*, vol. 27, pp. 1093–1112, 2021.

[4] H. Bi, W. L. Shang, Y. Chen, K. Wang, Q. Yu, and Y. Sui, "GIS aided sustainable urban road management with a unifying queueing and neural network model," *Applied Energy*, vol. 291, 2021.

[5] P. Mo, A. D'Ariano, L. Yang et al., "An exact method for the integrated optimization of subway lines operation strategies with asymmetric passenger demand and operating costs," *Transportation Research Part B: Methodological*, vol. 149, pp. 283–321, 2021.

[6] D. Pamucar, M. Deveci, F. Canıtez, and V. Lukovac, "Selecting an airport ground access mode using novel fuzzy LBWA-WASPAS-H decision making model," *Engineering Applications of Artificial Intelligence*, vol. 93, Article ID 103703, 2020.

[7] J. Yin, A. D'Ariano, Y. Wang, L. Yang, and T. Tang, "Timetable coordination in a rail transit network with time-dependent passenger demand," *European Journal of Operational Research*, vol. 295, no. 1, pp. 183–202, 2021.

[8] P. Mo, L. Yang, A. D'Ariano, J. Yin, Y. Yao, and Z. Gao, "Energy-efficient train scheduling and rolling stock circulation planning in a metro line: a linear programming approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3621–3633, 2020.

[9] D. Hörcher, D. J. Graham, and R. J. Anderson, "Crowding cost estimation with large scale smart card and vehicle location data," *Transportation Research Part B: Methodological*, vol. 95, pp. 105–125, 2017.

[10] K. M. Kim, S.-P. Hong, S.-J. Ko, and D. Kim, "Does crowding affect the path choice of metro passengers?" *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 292–304, 2015.

[11] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "A probabilistic Passenger-to-Train Assignment Model based on automated data," *Transportation Research Part B: Methodological*, vol. 104, pp. 522–542, 2017.

[12] B. Si, L. Fu, J. Liu, S. Shiravi, and Z. Gao, "A multi-class transit assignment model for estimating transit passenger flows-a case study of Beijing subway network," *Journal of Advanced Transportation*, vol. 50, no. 1, pp. 50–68, 2016.

[13] L. Sun, Y. Lu, J. G. Jin et al., "An integrated Bayesian approach for passenger flow assignment in metro networks," *Transportation Research Part C: Emerging Technologies*, vol. 52, pp. 116–131, 2015.

[14] B. Si, Z. He, D. Liu, X. Yang, and Z. Gao, "Train operation diagram-based equilibrium model for an urban rail transit network with transfer constraints," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 11, Article ID 04020126, 2020.

[15] P. Shang, R. Li, J. Guo, K. Xian, and X. Zhou, "Integrating Lagrangian and Eulerian observations for passenger flow state estimation in an urban rail transit network: a space-time-state hyper network-based assignment approach," *Transportation Research Part B: Methodological*, vol. 121, pp. 135–167, 2019.

[16] S. Raveau, J. C. Muñoz, and L. de Grange, "A topological route choice model for metro," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 2, pp. 138–147, 2011.

[17] S. Raveau, Z. Guo, J. C. Muñoz, and N. H. M. Wilson, "A behavioural comparison of route choice on metro networks:

time, transfers, crowding, topology and socio-demographics," *Transportation Research Part A: Policy and Practice*, vol. 66, pp. 185–195, 2014.

[18] B. Si, M. Zhong, J. Liu, Z. Gao, and J. Wu, "Development of a transfer-cost-based logit assignment model for the Beijing rail transit network using automated fare collection data," *Journal of Advanced Transportation*, vol. 47, no. 3, 2013.

[19] Y. Zhang, E. Yao, K. Zheng, and H. Xu, "Metro passenger's path choice model estimation with travel time correlations derived from smart card data," *Transportation Planning and Technology*, vol. 43, no. 2, pp. 141–157, 2020.

[20] C. Monterola, E. F. Legara, D. Pan, K. K. Lee, and G. G. Hung, "Non-invasive procedure to probe the route choices of commuters in rail transit systems," *Procedia Computer Science*, vol. 80, pp. 2387–2391, 2016.

[21] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Systems with Applications*, vol. 95, pp. 324–332, 2018.

[22] F. Zhou and R.-h. Xu, "Model of passenger flow assignment for urban rail transit based on entry and exit time constraints," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2284, no. 1, pp. 57–61, 2012.

[23] J. Zhao, F. Zhang, L. Tu et al., "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 790–801, 2017.

[24] W. Li, Q. Luo, Q. Cai, and X. Zhang, "Using smart card data trimmed by train schedule to analyze metro passenger route choice with synchronous clustering," *Journal of Advanced Transportation*, 2018.

[25] J. Wu, Y. Qu, H. Sun, H. Yin, X. Yan, and J. Zhao, "Data-driven model for passenger route choice in urban metro network," *Physica A: Statistical Mechanics and its Applications*, vol. 524, pp. 787–798, 2019.

[26] S. Das Mukherjee, "Application of transformer-based language models to detect hate speech in social media," *Journal of Computational and Cognitive*, pp. 11–20, 2021.

[27] M. Rahbar, M. Hickman, M. Mesbah, and A. Tavassoli, "Calibrating a bayesian transit assignment model using smart card data," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2018.

[28] C. Yu, H. Li, X. Xu, and J. Liu, "Data-driven approach for solving the route choice problem with traveling backward behavior in congested metro systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, Article ID 102037, 2020.

[29] Y. Zhang, E. Yao, J. Zhang, and K. Zheng, "Estimating metro passengers' path choices by combining self-reported revealed preference and smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 76–89, 2018.

[30] E. Jenelius and H. N. Koutsopoulos, "Travel time estimation for urban road networks using low frequency probe vehicle data," *Transportation Research Part B: Methodological*, vol. 53, pp. 64–81, 2013.

[31] N. Uno, F. Kurauchi, H. Tamura, and Y. Iida, "Using bus probe data for analysis of travel time variability," *Journal of Intelligent Transportation Systems*, vol. 13, no. 1, pp. 2–15, 2009.

[32] L. Zhang, G. Zeng, D. Li, H.-J. Huang, H. E. Stanley, and S. Havlin, "Scale-free resilience of real traffic jams," *Proceedings of the National Academy of Sciences*, vol. 116, no. 18, pp. 8673–8678, 2019.

[33] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Comparison of methods of computing lognormal sum distributions and outages for digital wireless applications," in *Proceedings of the Conference Record - International Conference on Communications*, pp. 1270–1275, IEEE, June 1994, Stockholm, Sweden, 1994.

[34] M. Deveci, D. Pamucar, and I. Gokasar, "Fuzzy Power Heronian function based CoCoSo method for the advantage prioritization of autonomous vehicles in real-time traffic management," *Sustainable Cities and Society*, vol. 69, Article ID 102846, 2021.

[35] B. Y. Chen, Q. Li, and W. H. K. Lam, "Finding the k reliable shortest paths under travel time uncertainty," *Transportation Research Part B: Methodological*, vol. 94, pp. 189–203, 2016.

[36] D. Sirbiladze, B. Midodashvili, L. Midodashvili, and D. Siprashvili, "About one representation-interpeter of a monotone measure," *Journal of Computational and Cognitive*, pp. 1–8, 2021.

[37] V. Simic, I. Gokasar, M. Deveci, and M. Isik, "Fermatean fuzzy Group decision-making based CODAS approach for taxation of public transit investments," *IEEE Transactions on Engineering Management*, pp. 1–16, 2021.

[38] W. L. Shang, Y. Chen, H. Bi, H. Zhang, C. Ma, and W. Y. Ochieng, "Statistical characteristics and community analysis of urban road networks," *Complexity*, vol. 2020, Article ID 6025821, 2020.

[39] W.-L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: a big-data analysis," *Applied Energy*, vol. 285, Article ID 116429, 2021.